Spring 6-2020

Second Language Acquisition and Standardized Assessment Practices

Hal Scott Schmeisser
DePaul University

DePaul University

College of Education

SECOND LANGUAGE ACQUISITION AND STANDARDIZED ASSESSMENT PRACTICES

A Dissertation in Education
with a Concentration in Educational Leadership

by

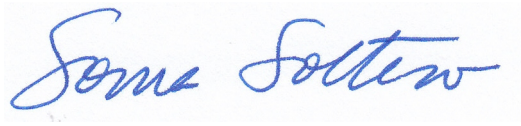
Hal S. Schmeisser

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Education

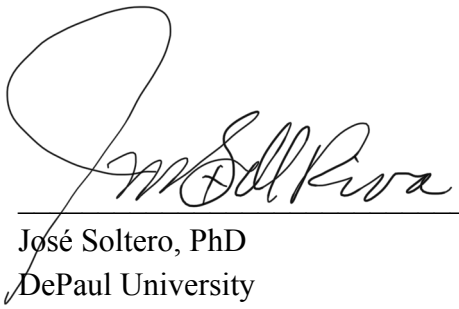
June 2020

We approve the dissertation of Hal S. Schmeisser.



Sonia Soltero, PhD
DePaul University
Professor
Leadership Language and Curriculum (Chair)
Chair of Dissertation Committee

Date



5/8/2020

José Soltero, PhD
DePaul University
Associate Professor
Sociology
Dissertation Committee Member

Date

Sung Park-Johnson, PhD
DePaul University
Assistant Professor and Director
Leadership Language and Curriculum
Dissertation Committee Member

Date

Certification of Authorship

I certify that I am the sole author of this dissertation. Any assistance received in the preparation of this dissertation has been acknowledged and disclosed within it. Any sources utilized, including the use of data, ideas and words, those quoted directly or paraphrased, have been cited. I certify that I have prepared this dissertation according program guidelines, as directed.

Author Signature  Date 4/23/2020

ABSTRACT

Throughout the history of the United States, tests have been utilized as a means to exclude and marginalize immigrant and minority populations, and in the absence of a national language policy have served as de facto language policy. English Learners (ELs) have been among the fastest-growing student populations in the United States, comprising a heterogeneous group with different cultural and linguistic backgrounds. Research findings concerning the assessment of ELs highlight a significant difference in the academic performance of these students from that of their native English speaking classmates. ELs face a substantial hurdle in learning both English and academic content in English at the same time.

The purpose of this quasi-experimental correlational design quantitative study was to examine the relationship between ELs' second language proficiency and achievement on federally mandated standardized content-area assessments in ELA and mathematics. This study sought to address the overarching issues concerning the processes of second language development and assessment of ELs within the context of public education in the United States.

This study employed various statistical tests - linear correlation, Chi-Square test of independence, and linear regression models - to measure the degree of association and strength of relationship amongst the variables to test the hypotheses and answer the research questions. The participants for this study included 119 third, fourth, and fifth grade ELs in a small, yet significantly diverse, suburban Illinois school district during the 2017-2018 academic school year.

The results of this study empirically demonstrated that the PARCC assessment for both ELA and mathematics is a linguistically laden challenge for ELs in a way that it is not for students whose primary language is English. The results clearly elucidate the relationship between English language proficiency and achievement on PARCC, a federally mandated performance-based content-area assessment.

Furthermore, this study provides specific information and insight regarding the validity and reliability of standardized assessment practices within the State of Illinois at a moment of decision and opportunity. There are profound societal impacts to this work that extend well beyond performance on a standardized assessment. The equity in assessment and educational experience for ELs is critical to our work in schools of providing all students with the knowledge, understandings, strategies, and confidence to achieve success in a self-determined future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF GRAPHS

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to extend my deepest gratitude and appreciation to my family. Each of you gives my life meaning and fulfillment in ways one could only dream of. To my daughter Liesel, you are our everything and inspire beyond expression. Your inquisitive nature and relentless persistence are qualities I find admirable and strive to mirror. To my wife, best friend, and the most outstanding educator I know, Alicia - your encouragement, confidence, and unwavering support has helped me to grow into the person I am. I could not possibly imagine my life without you and am so grateful for our partnership each and every day. To my in-laws, Ed and Alice, I simply cannot thank you enough for everything you do for our family. Without your love, support, and kindness none of this would be possible. You add such light to our lives. To my dad and mom, Steve and Kathy, thank you for the foundation that you provided me in the journey of life, and for the terrible sense of humor that has made it so fun.

Thank you Dr. Sonia Soltero for your willingness to take on this project with me, and for your endless guidance and support. Your ability to translate this work into meaningful and actionable outcomes for kids fills me with inspiration and lifts me with hope. Thank you Dr. José Soltero for your patience, guidance, and insight into quantitative statistics and research. You have shaped a truly novice mathematician into a respectable researcher. Thank you Dr. Park-Johnson for your support and perspicacity. Your expertise within the field has contributed greatly to this research and pushed my thinking as both a teacher and leader.

Lastly, I want to thank the students who inspired this research. It is your courage that led me to this topic, as well as to better understanding the very real and significant impact of this work. It is your future, still to be written, that is at the heart of this endeavor!

CHAPTER ONE. INTRODUCTION.

This research topic was selected from the infinite possibilities of dissertation theses for its situation at the intersectionality of race, culture, class, and identity issues. It is my belief that addressing the equity of opportunity in the education of English Learners (ELs) must be a top priority for us as educators and engaged members of a democratic society. This is a group of children whose abilities have for too long been undervalued and discounted, while having their needs poorly served and underrepresented.

In thirteen years of experience within the field of public education, encompassing both teaching and leadership roles, I have had the opportunity to serve four different school communities. Each with its own unique strengths and challenges in meeting the needs of children, particularly those developing English as another language. Consistent amongst each of these communities and their school districts, is the struggle to effectively support ELs and their families in robust and meaningful ways. While the EL population of each environment is distinct - racially, ethnically, culturally, linguistically, economically, and spiritually - the needs of these children have been systematically underserved and overlooked. This, despite positive intentions and pockets of excellence and care. As a teacher, and now as an administrator, I found myself living in larger systems that did not consider the very specific and complex needs of ELs. In each district the richness and complexity of growing and developing young people was traded for data sets embedded in spreadsheets, graphs, and tables. The challenge of truly getting to know each child as a learner and young person while helping them to grow from their current state of being was swapped out for vague visions and strategic goals aligned to everything but the needs of our most vulnerable students. Challenging this status quo and directly confronting

the linguistic, academic, and emotional needs, as well as, assessment realities of ELs typically garnered one of three responses. These I believe are symptomatic of education as a whole. The initial response is often one of acceptance, that there exist broader systems at play outside of our control. This "render unto Caesar" scenario fails to acknowledge the role of administrators, teachers, and all educational actors in negotiating and acting upon policy at levels throughout the system. The next response is simply to avoid the matter and its underlying complexities. This line of thinking follows that to not highlight and discuss the issue renders it invisible and thus not an issue at all. I believe the third response, reframing the narrative, is the most common. In reframing, educators cast the unique needs of ELs within other demographic terms, like race, ethnicity, and socioeconomic status. Though perhaps well intended, this response fails to recognize, understand, and grapple with the multidimensionality of the matter; it leads to fragmentary solutions based on a limited understanding of identity.

Upon this recognition, I started to explore the concept of new managerialism, seeking to find an understanding for how the machinations of markets so strongly found their way into the systems, structures, and discourse of education. I also expanded my studies on cultural validity and other extensive equity issues within education. With experiences at the individual student, classroom, school, and district level, I began to more clearly understand and connect the broader layers of legislation, policy, and judicial rulings that impacted educational actors throughout the system. As I continued to read, reflect, teach, and lead I kept returning to the larger framework of high stakes assessment and accountability that encompasses all of education. Though we rarely discuss it as such, this system of standardized assessment reigns over the entirety of education. It determines which schools and students are labeled successful, and which are

defined as failing. It specifies the allocation of financial resources and supports, even deciding which communities are capable of self-governing their own schools. Often deemed "low stakes" and non-immediate for individual students, these broader consequences shape the practices and policies experienced every day by children. The fallout of high-stakes assessments impacts the educational models, programming, curriculum, classroom make-up, and instructional practices of all students. This holds even more true for groups of students deemed "unsuccessful" and "at-risk," disproportionately students who are culturally and linguistically diverse. The influence of this set-up is real and experienced by both children and adults daily. Standardized assessment and the accountability systems that accompany it play an oversized role within the classrooms, hallways, and boardrooms of our state's schools. It is my sincere hope that this dissertation helps to move this conversation forward in authentic, specific, and above all actionable ways.

To borrow inspiration from Dr. Laurie Olsen's *Reparable Harm* (2010), this dissertation is a summons to put our collective voices and actions towards creating assessment systems and accountability structures that serve all children - "to recognize, respond, and to repair the harm that has been exacerbated by lapses in state, district, and school policies and practices. It is reparable harm, it is preventable harm, and it is wholly in our power to change" (p. iii). The broader question looming beyond this research is: now that we know better, do we have the courage to do better?

Statement of the Problem

Throughout the history of the United States, literacy tests have been utilized as a means to exclude and marginalize immigrant populations, and in the absence of a national language policy have served as de facto language policy (Crawford, 2000; Menken, 2008; Wiley & Wright, 2004).

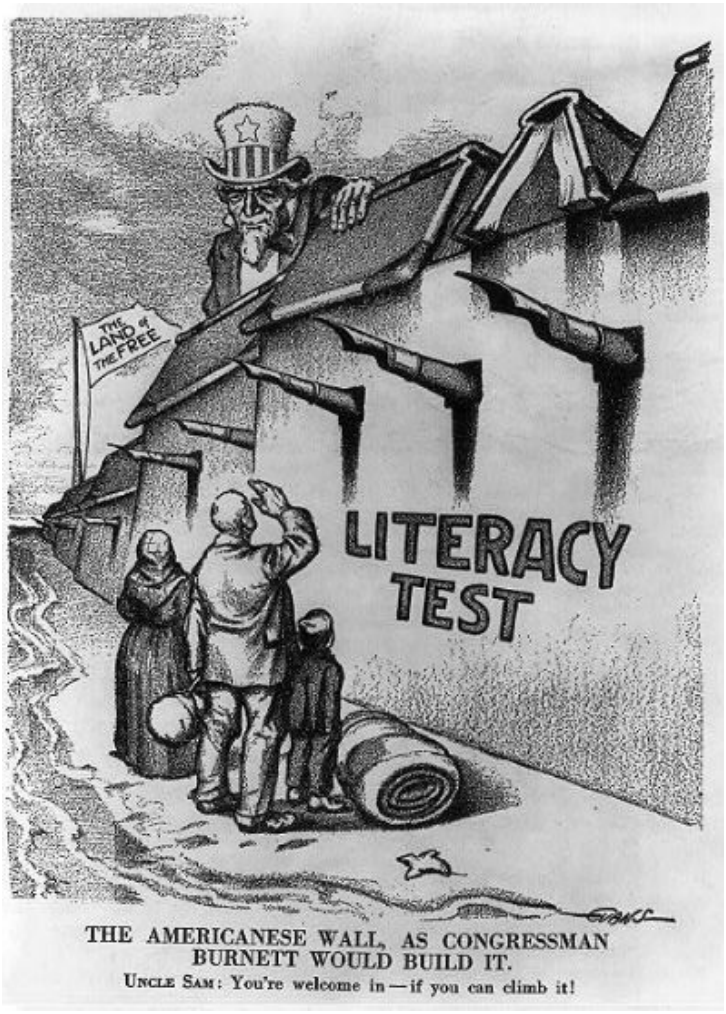


Figure 1. "The Americanese wall - as Congressman [John Lawson] Burnett would have build it," by Raymond O. Evans, 1916, *Library of Congress, Prints and Photographs Division.* Retrieved November 1, 2018, from https://www.loc.gov/item/2006681433/?loclr=blogtea

The purpose of this study is to examine the theory and research that addresses the second language acquisition and assessment of ELs within the context of public education in the United States. The study seeks to address the overarching issue of the relationship between an ELs' second language proficiency and their performance on federally mandated standardized content-area assessments.

It is important to contextualize this study within a rapidly changing and increasingly diverse United States population. Throughout the twenty-first century, ELs have been among the fastest-growing populations in the nation's schools, comprising ten percent of the student population nationwide (United States Department of Education, 2017). The National Center for Education Statistics places this number at 4.6 million students for the 2014-15 school year (Center for Applied Linguistics, 2018). Given varying standards and processes across states for identifying ELs, and a changing socio-political landscape, overall counts vary. According to the United States Census Bureau 2016 American Community Survey, 5% of US children ages 5-17 were identified as Limited English Proficient (LEP) or EL, while data submitted by individual states to the federal government put the percentage of ELs in kindergarten through twelfth grade at ten percent (United States Department of Education, 2017; National Center for Education Statistics, 2018).

ELs comprise a heterogeneous group of students with different cultural and linguistic backgrounds (Center for Applied Linguistics, 2017; Migration Policy Institute, 2018; United States Department of Education, 2017). The Annie E. Casey Foundation (2018) estimated that, in 2016, 22% of children in the United States spoke a language other than English (LOTE) at home (Center for Applied Linguistics, 2018). As of 2013, 85% of pre-kindergarten through fifth

grade students were born in the United States, while 62% of sixth through twelfth grade students were born in the United States (Zong & Batalova, 2015). As measured through school demographics, the United States is more diverse than ever before. Public school enrollment in 2017 shows that, for the first time ever, majority-minority children constituted 51% of the student population, students from low income households also made up 51% of the population (Center for Applied Linguistics, 2018). The majority of ELs were born in the United States and are legal citizens (Sanchez, 2017).

In Illinois in 2016, approximately 1,783,000 individuals were foreign-born, composing 14% of the state's population. This make-up roughly mirrors the broader United States population, with 13.5% of individuals nationwide being of foreign-birth (Migration Policy Institute, 2018). The beginning of the twenty-first century marked a steep decline in the growth rate of the foreign-born population in Illinois - dropping from 61% during the 1990's, to 17% between 2000-2016 (Migration Policy Institute, 2018). Even with this trend, the state's immigrant population continues to increase more rapidly than its native-born population (Migration Policy Institute, 2018). It should be noted as a related aside that Puerto Ricans are US citizens, also serve ELs within their education system, and are technically not "foreign born." Age group and low-income trends in Illinois closely mirror broader trends in the United States (Migration Policy Institute 2018; MPI Data Hub 2018). According to the Illinois State Board of Education's *2017 Annual Report*, during the 2016-17 school year 10.7% of students within the state's K-12 system were classified as ELs. This constitutes 205,585 children, demonstrating a gradual increase of over 3% for the state's EL population in the past decade (Illinois State Board of Education, 2018). In Illinois, more than three quarters of school-aged children, reported as

ELs in census data, were born in the United States. This is slightly above the national average native-born LEP rate of 71% (Migration Policy Institute, 2018). Spanish is the majority native language of ELs both in the United States and Illinois. 76.7% of ELs in Illinois identify Spanish as their native language; Arabic (3.67%) is a distant second, with Polish (2.93%), Urdu (1.56%), and Pilipino/Tagalog (0.93%) rounding out the top five (Illinois State Board of Education, 2018).

Federal education legislation of the past two decades has demonstrated a significant shift towards emphasizing standardized assessments within a system of high-stakes accountability (Abedi, 2003; Menken, 2009; Hopkins, Thompson, Linquanti, Hakuta, & August, 2013). While these policies have impacted all students, there has been a broader and disproportionately greater negative impact on ELs and the school communities that serve them (Ho, 2008; Kortez, 2008; Linn, 2005, 2008). While some researchers note the positivity of this legislation in highlighting a previously invisible population of students (Hopkins, Thompson, Linquanti, Hakuta, & August, 2013), others address the significant overreach of the federal government into the local issue of education (Crawford; 2002; Menken, 2008, 2009; Wiley & Wright, 2004). Despite these varying interpretations, consistent across the research is the emphasis of federal legislation on standardized assessment within a system of high-stakes accountability and decision-making. Additionally, the *Every Student Succeeds Act* (ESSA) mandates that ELs must take both academic content-area and language proficiency assessments (Lyons & Dadey, 2017).

The rationale for including ELs in these mandated English language proficiency and grade level content-area assessments and overarching systems of accountability is to ensure their educational progress relative to English language development and grade level content-area standards (Abedi, 2003; Menken, 2009). However, the most salient factor concerning the

validity of assessment for ELs is that these standardized assessments are not designed, nor implemented, with the EL student population in mind (Abedi, 2011; Kopriva, 2000; Solano-Flores, 2011; Trumbull et al, 2011). Research findings concerning the assessment of ELs highlight a significant difference in the academic performance of these students from that of their native English speaking classmates (Abedi, 2004, 2006; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Solano-Flores & Trumbull, 2003). ELs face a significant hurdle in learning both English and academic content in English at the same time. This challenge is intensified on grade level content-area assessments that are densely worded with academic vocabulary and ripe with linguistic complexity (Abedi, 2002, 2006, 2011; Solano-Flores, 2008). Additionally, ELs are often supported on these assessments with ineffective accommodations that were not designed to meet their specific linguistic needs (Abedi, Hofstetter, & Lord, 2004; Abedi, 2007, 2011; Lara & Chia, 2011). Research findings indicate that it takes five to seven years for most ELs to gain mastery of academic English at a level comparable to native English speaking peers (Collier & Thomas 1989, 2005, 2009; Cummins, 1981, 2012; Hakuta, Butler, Witt, 2000). Hence, there exists substantial misalignment between the research, aims, policies, and practices surrounding the assessment and educational experiences of ELs.

Unintended consequences of this misalignment and system of high-stakes testing for ELs include: increased dropout rates, decreased graduation rates, higher rates of younger students taking the General Education Diploma (GED) instead of completing high school, increased numbers of low performing students being held back in grades before testing, and higher numbers of suspensions and expulsions of low performing students immediately prior to testing (Amrein & Berliner, 2002; Menken, 2008). Additionally, there is a more personalized impact of

such a deficit-based approach to education that serves to further separate students from their linguistic, cultural, and community-based identities (Valenzuela, 2002, 2005; Menken, 2008). Increased accountability, the kind that provides the frameworks of federal legislation like *No Child Left Behind* (NCLB) and *Every Student Succeeds Act* (ESSA), "reduce the quality and quantity of education offered, and has the most damaging effects on poor and minority youth" (Menken, 2009, p. 98). While native language assessments can offer advantages for some students to more accurately confer their academic competency, research is clear about the need to align the language of assessment with that of instruction (Abedi, 2011; Menken, 2009). There also exists additional challenges and equity issues in determining which languages to offer, as well as, creating parity between assessments. Taken together, the combination of invalid assessment practices which measure irrelevant constructs, ineffective accommodations, and a culture of high stakes accountability create a unique educational context for ELs, as well as, the districts, schools, classrooms, and communities which serve them.

The growing population of ELs within the United States operating within a system of standardized assessment and high-stakes accountability, leads to an increased sense of urgency in addressing these issues, remediating achievement gaps, and better meeting the educational needs of this unique student population. There exists a need to further address the contradiction between the stated aims and actual impact of assessments at the federal, state, district, school, and classroom levels. Investigating current assessment policies and practices, while seeking to create a more valid and equitable system of assessment for ELs is important as these tests currently drive the instructional practices, decision-making, and educational opportunities of ELs in the United States (Menken, 2009). With further inaction, our current educational system will

continue to reproduce its inequities; there is a profound responsibility to examining our assessment structures, systems, practices, and beliefs in order to improve and create a more equitable educational experience for ELs.

Purpose of the Study

The purpose of this study is to examine the relationship between EL's second language proficiency and their academic achievement on federally mandated standardized content-area assessments. The independent variable, English language proficiency, is defined as a student's overall scale score and accompanying proficiency level on the ACCESS for ELLs 2.0 assessment. The ACCESS assessment is used throughout 40 states, including Illinois, that make up the WIDA Consortium to measure an EL's level of language proficiency in English. This computer-based standardized test assesses four language domains - reading, writing, speaking, and listening - providing a scale score (100-600) and accompanying proficiency level (1.0-6.0). These levels span from the beginning stages of English language development (entering and emerging) towards middle (developing and expanding) and later (bridging and reaching) stages of proficiency. The dependent variable, achievement on grade level content-area standardized assessment, is defined as a student's overall scale scores and accompanying proficiency levels on the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment for both ELA and mathematics. PARCC is a next generation performance-based assessment initially developed as part of the federal government's *Race to the Top* competitive grant. It is a computer-based standardized test that assesses the subdomains of literary text, informational text, and vocabulary for reading, and writing expression and knowledge and use of language conventions for writing. The mathematical subdomains assessed are major content, additional

and supporting content, expressing mathematical reasoning, and modeling and application. PARCC performance is measured by a scale score for both ELA and mathematics, ranging from 650 to 850. This scale score is then associated with one of five levels - Did Not Meet, Partially Met, Approached, Met, Exceeded - used to determine a student's proficiency towards achieving state content standards for a particular grade level. The relationship between the ACCESS and PARCC scores for a small, yet diverse, group of third through fifth grade ELs will be investigated in this quasi-experimental correlational design quantitative study. Furthermore, this study will use multivariate regression analysis to explore the impact of English language proficiency relative to other student demographic variables (ethnicity, socioeconomic status). The study uses the lens of cultural validity as defined by Solano-Flores and Nelson-Barber (2001) to explore the extent to which current assessment practices move us closer to the goal of ensuring equitable educational opportunity for ELs. The concept of cultural validity provides the opportunity to examine the socio-cultural factors that shape student thinking and impact assessment outcomes for ELs, as well as, investigate large-scale standardized assessment practices from a cultural perspective. It is hoped that this research will provide another tool to the field for objectively examining the extent to which testing practices are sensitive to the issues of language and culture. The results of this study seek to inform policymakers and educators at the state, district, school, and classroom level as to the most equitable, valid, and just assessment practices and policies for ELs and the school communities that serve them.

Research Questions and Hypotheses

The overall research question guiding this study is:

How does an ELs' second language proficiency influence his/her achievement on federally mandated standardized content-area assessments?

The subsequent research questions and hypotheses for this study were as follows:

Research question 1. What is the relationship between English language proficiency and achievement on grade level content-area standardized assessments in ELA and mathematics?

Hypothesis 1. The more advanced the student's level of language proficiency in English, the higher their reading achievement.

Hypothesis 2. The more advanced the student's level of language proficiency in English, the higher their mathematics achievement.

Research question 2. To what extent does English language proficiency in reading, writing, speaking, and listening influence achievement on grade level content-area standardized assessments in ELA and mathematics?

Hypothesis 3. A student's English language proficiency in reading and writing will impact their reading achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.

Hypothesis 4. A student's English language proficiency in reading and writing will impact their mathematics achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.

Research question 3. What is the threshold English language proficiency for ELs to meet/exceed standards on federally mandated standardized assessments in ELA and mathematics?

Hypothesis 5. According to Illinois school code ELs are defined as proficient at a composite ACCESS score of 4.8 or beyond, and should thus be able to demonstrate proficiency on grade level content-area assessments.

Research question 4. What is the impact of other student demographic factors - race/ethnicity, socioeconomic status - in relation to English language proficiency for the achievement of ELs on federally mandated standardized assessments in ELA and mathematics?

Hypothesis 6. Controlling for other variables, ELs identifying with a race/ethnicity of White will achieve higher levels of reading and mathematics proficiency.

Hypothesis 7. Controlling for other variables, ELs with full pay lunch status will achieve higher levels of reading and mathematics proficiency.

Significance of the Study

Due to the increasing population of ELs within the United States and its public school system, as well as, a system of increased accountability rooted in high-stakes assessment, it is critical for researchers and educators to identify the most equitable and valid assessment practices and policies to ensure the unique needs of ELs are being met. A study emphasizing standardized assessment within a lens of socio-cultural and linguistic validity in the education of ELs has particular significance at this time.

This study seeks to address a current gap in the literature by bringing together different research domains within a real world context, and more fully investigating the relationship between second language proficiency and performance on state-mandated assessments, particularly with digital performance-based assessments. A quantitative orientation was selected in order to provide an analysis in the language of new managerialism comprehensible to current policy-makers and many educational reformers. According to Lynch, Grummell, and Devine (2012), "within education, new managerialism redefined what counts as knowledge, who are the bearers of such knowledge and who is empowered to act - all within a legitimate framework of public choice and market accountability" (p. 4). Thus the intention of such an orientation was to utilize this language of power to posit transformations of our educational system. This research focuses on the need to provide a more equitable and valid assessment and educational experience for ELs.

Given the ongoing emphasis of standardized assessment as a driver for accountability and school improvement efforts both at the national and state level, there is particular relevance and significance to this study. This study is timely as the State of Illinois reviews its annual grade level content-area assessment for ELA and mathematics, and considers a shift towards a new assessment for the 2020-21 school year. Additionally, the current debate over immigration and political rhetoric in the United States represents a pivotal moment for addressing the educational opportunities and inequities for those children learning to speak English as a second language. There are profound societal impacts to this work that extend well beyond performance on a standardized assessment. The equity in assessment and educational experience for ELs is critical

to our work in schools of providing all students with the knowledge, understandings, strategies, and confidence to achieve success in a self-determined future.

Definition of Terms

The terms used in the study are defined as follows:

- English learner (EL): prior to the passage of the *Every Student Succeeds Act* in December 2015, the term Limited English Proficient (LEP) was used by federal and state agencies to describe ELs. As defined by the law, the term "English learner," when used with respect to an individual, means an individual (ESEA Section 8101(20)) —

  > (A) who is aged 3 through 21;
  > (B) who is enrolled or preparing to enroll in an elementary school or secondary school;
  > (C)(i) who was not born in the United States or whose native language is a language other than English;
  > (ii)(I) who is a Native American or Alaska Native, or a native resident of the outlying areas; and (II) who comes from an environment where a language other than English has had a significant impact on the individual's level of English language proficiency; or
  > (iii) who is migratory, whose native language is a language other than English, and who comes from an environment where a language other than English is dominant; and
  > (D) whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual —
  > (i) the ability to meet the challenging State academic standards;
  > (ii) the ability to successfully achieve in classrooms where the language of instruction is English; or
  > (iii) the opportunity to participate fully in society.

  Due to its deficit-based approach the term LEP has fallen out of use in recent years; however, LEP has precise meaning in court decisions, federal and state laws (Crawford, 2004, p. xiv).

- English as a second language (ESL): as defined by the Illinois School Code, the term means (Implementing Article 14C and authorized by Section 2-3.39(1) of the School Code [105 ILCS 5/Art. 14C and 2-3.39(1)]) -

  > "specialized instruction designed to assist students whose home language is other than English in attaining English language proficiency. ESL instruction includes skills development in listening, speaking, reading and writing. (ESL is not to be confused with English language arts as taught to students whose home language is English."

- English Language Development (ELD) Standards: as defined by the Illinois School Code, the term relative to the age group examined in this study - children over the age of five and a half - means (Implementing Article 14C and authorized by Section 2-3.39(1) of the School Code [105 ILCS 5/Art. 14C and 2-3.39(1)]) -

  > "the: "2012 Amplification of English Language Development Standards Kindergarten-Grade 12" (2012) for students in kindergarten and grades 1 through 12 published by the Board of Regents of the University of Wisconsin System on behalf of the World-class Instructional Design and Assessment (WIDA) Consortium, Wisconsin Center for Education Research (WCER), University of Wisconsin-Madison, 1025 West Johnson Street, MD#23, Madison WI 53706 and posted at http://wida.us/standards/eld.aspx (no later amendments to or editions of these standards are incorporated)"

Limitations

There exist some potential limitations to this study, which will be further discussed and expanded upon in the methodology section to follow. First, is the relatively small sample size (n = 119). The sample size is such that it is difficult to generalize the findings to a larger population of students, particularly for a group of ELs that may be more linguistically homogeneous. Another limitation is the pre-existing differences amongst the EL student population used in the

study. Each child is unique, having their own personal and educational experiences. A combination of these external factors poses some threat to the validity of this study. Similarly, the study has limitations with respect to the generalizability of the racial and ethnic identities of the student population. Additionally, lunch status was used as a proxy for the socioeconomic levels of families, with no readily available information on parent's educational attainment, thus contributing further limitations to the study. Finally, the use of standardized assessments itself to measure language proficiency and knowledge more generally is an inherent limitation. As this study and a wealth of supporting research detail, language is complex, dynamic, and socially constructed. To attempt to define one's linguistic capabilities using a standardized measure will surely miss the mark on the range of possibilities for any individual. The same can be said for the use of standardized assessments in assessing human knowledge and expected learning. Thus, and perhaps ironically, the use of standardized assessment data to analyze the impact of such data is in itself a limitation to the study.

Nature and Order of the Presentation for the Study

Chapter two provides a review of the literature in order to frame the knowledge base, relevant theories, and existing research for the broader study. The literature review discusses: (a) historical and contemporary context of standardized assessment and language policy within the United States; (b) theoretical frameworks of second language acquisition; (c) stages, processes, and duration of second language acquisition for school age children; (d) divide between school practices and current second language acquisition research; (e) critical issues relevant to ELs and assessment, including the cultural and linguistic validity of standardized assessments.

Chapter three discusses the research methodology, explaining the quasi-experimental correlational design for the proposed quantitative study. It includes a restatement of the research purpose, questions, and hypotheses, as well as, an overview of assessments in Illinois and examination of specific testing items within the PARCC assessment. The chapter also describes the site selection, research design, data collection and analysis, validity, ethical issues, and limitations of the proposed study. The chapter concludes by restating the goals of this research in striving to contribute to a broader scholarly discussion.

Chapter four provides an overview of the data and analysis of the results relating to each of the research questions and underlying hypotheses. The four research questions and seven corresponding hypotheses are utilized as a framework for presenting the findings of this study. The data collection procedures and results of the statistical tests applied for each question are reviewed, and are followed by an analysis of the data for each research question. The chapter concludes with a summary of the findings.

Chapter five marks the conclusion and begins by providing a summary of the study. The findings of the study are then related to each of the research questions and accompanying hypotheses, as well as, linked to theories and research that contribute to an overarching theoretical framework. The implications of the study and suggestions for future research are discussed. The chapter ends with a conclusion of the study and its intended impact.

CHAPTER TWO. LITERATURE REVIEW.

Introduction

The purpose of this chapter is to examine the theory and research that addresses the second language development and assessment of English Learners (ELs) within the context of public education in the United States. The review seeks to address the following question: is there a relationship between ELs' second language proficiency and their performance on state standardized tests? This review explores and analyzes the findings of researchers and educational theorists within the fields of second language acquisition and educational assessment. The rationale for this review and proposed study is to move toward a more comprehensive exploration and understanding of these issues for ELs and their families, as well as, the schools and communities that serve them.

The specific aims of this literature review are to: (a) briefly overview the historical and contemporary context of standardized assessment and language policy within the United States; (b) analyze the theoretical frameworks of second language acquisition; (c) describe the stages, processes, and duration of second language acquisition for school age children; (d) analyze the divide between school practices and current second language acquisition research; (e) examine critical issues relevant to ELs and assessment, including the cultural and linguistic validity of standardized assessments. This section will conclude with an overview regarding the validity and equity of current high-stakes standardized assessment practices for ELs and rationale for this study.

There is an extensive amount of literature, spanning the past four decades, regarding the length of time, processes, and stages of second language acquisition (Collier 1987, 1989; Collier

& Thomas, 1995, 2007, 2009 ; Cummins, 1981, 2012; Hakuta, Butler, & Witt, 2000). Much of this research builds upon the body of work that preceded it and is centered around the groundbreaking ideas of central theorists and researchers (e.g. Cummins, Krashen, Collier & Thomas). There also exist substantial literature on the tensions of ELs and standardized assessments, particularly with respect to addressing language barriers, psychometric concerns, and accommodations (Abedi, 2002, 2006, 2011; Hakuta, Butler, & Witt, 2000; Solano-Flores, 2001, 2008, 2011). There is a robust field of scholarship exploring the high-stakes assessment and accountability structures for ELs (Crawford, 1992, 1999, 2002; Gàndara, 2015; Menken, 2000, 2008, 2009; Wright, 2010, 2017). A current gap in the literature lies with respect to bringing these different research domains together within a real world context, and more fully exploring the relationship between second language proficiency and performance on state-mandated assessments, particularly with computerized performance based assessments like PARCC.

Mandated Identification of ELs in Illinois

Illinois School Code Article 14C outlines the requirements for districts to provide transitional bilingual education and provides financial assistance to help districts meet the additional costs of such programming, while also allowing the state to provide technical assistance and professional development to support this transitional bilingual programming (Illinois State Board of Education, 2019). In accordance with Article 14C of the School Code, 23 Illinois Administrative Code 228, establishes requirements for school districts' providing services to students in preschool through grade 12 who have been identified EL. As such, public school districts in Illinois must provide a Transitional Bilingual Education program for ELs,

except in attendance centers that meet flexibility requirements based upon enrollment numbers (Illinois State Board of Education, 2019). Illinois Administrative Code, Part 228, Section 228.15 requires school districts administer an individual language proficiency assessment to each student identified through the home language survey within 30 days of enrollment in the district (Illinois State Board of Education, 2019). Per the statute, English as a second language (ESL) means "specialized instruction designed to assist students whose home language is other than English in attaining English language proficiency. ESL instruction includes skills development in listening, speaking, reading and writing" (Illinois State Board of Education, 2019). The standards by which English language proficiency is measured are the "2012 Amplification of English Language Development Standards Kindergarten-Grade 12" created by the World-Class Instructional Design and Assessment (WIDA) Consortium in partnership with the University of Wisconsin-Madison. Additionally, the "English Language Proficiency Assessment" and "Prescribed Screening Instrument" outlined in the code are the WIDA assessment suite, including ACCESS 2.0 for ELs. For the purposes of this section, the terms "limited English proficient student" and "students with limited English proficiency", as used in Article 14C of the School Code, are understood to be ELs.

The home language survey is generated by the Illinois State Board of Education and available in English and forty other languages. Students whose families respond "yes" to one or both of the following questions, "Is a language other than English spoken in the home?" and "Does the student speak a language other than English?" are to be assessed for their English language proficiency to determine their eligibility for bilingual education services (Illinois State Board of Education, 2019).

Beginning in the 2017-18 school year the WIDA Screener assessment is the tool to be used for the purposes of identifying ELs the second semester of first grade through twelfth grade. The WIDA Screener assesses four primary language proficiencies: reading, writing, speaking, and listening. The WIDA Measure of Developing English Language (MODEL) assessment is utilized for students entering during the second semester of kindergarten or first semester of first grade, and assesses all four language proficiencies. Students entering during the first semester of kindergarten are only assessed utilizing the listening and speaking domains of WIDA MODEL. Pre-kindergarten students are screened for English proficiency using the Pre-IPT Oral English assessment published by Ballard & Tighe (Illinois State Board of Education, 2019). On each of these initial screening assessments, students receive a composite numeric score indicating their overall level of English language proficiency. This cut score, 5.0 overall & composite literacy 4.2 - reading/writing for the WIDA screener, and baseline performance level is then used to determine a student's eligibility for bilingual education and language support services (Illinois State Board of Education, 2019).

School districts are required to inform families in writing of the results of these screening assessments, as well as any recommendations for educational programming and services available to support the student in becoming English proficient. Parents not wanting their children to receive ESL or bilingual education may refuse these supports, but are required to formally document this refusal in a written statement. Parental refusal of services does not absolve the district from its responsibility to provide a meaningful education to the student, and still requires the student to be tested annually to determine their language proficiency (Illinois State Board of Education, 2019). The following section explores the history of language policy

and standardized assessment in the United States, presenting a historical and modern day context for the existence of such policies.

A Brief History of Language Policy and Standardized Assessment in the United States

The United States does not have a national official language; however, more than half of the states do have English as their official language. Illinois passed a law in 1923 making "American" its official state language, before amending the law in 1969 and changing it to "English." Though this law, and others similar to it, are not enforced; it does serve a broader symbolic purpose. Furthermore, this law runs contrary to legislation, like Article 14C, requiring bilingual education, as well as, other mandates requiring interpreters for medical services and court proceedings for citizens not speaking English as their primary language. Accordingly legislation, policies, judicial rulings, and everyday interpretations and practices all have an impact on the operation and parameters of teaching and learning (Menken 2000; 2008; 2009). Situating the literature within this broader context, offers greater perspective as to the shifting socio-political landscape impacting ELs educational experience.

Language Policies in Education

Language policies pertaining to education have shifted throughout the history of the United States, alternating between increasingly expansive and restrictive approaches to linguistic diversity (Menken, 2008). From colonial history through the present times, cultural and linguistic diversity has at times been tolerated, while restricted at others points (Crawford 1992; Kloss, 1977; Ricento, 1995). There exists an ebb and flow related to broader immigration patterns, political rhetoric, and social issues, amongst various other factors (Crawford, 1992, 1999; Kloss, 1977; Menken, 2008, 2009; Ricento, 1995).

Absent an explicit mention of language in the Constitution, the early history of the United States was a time of multilingualism in schools. Many schools during the colonial era taught, and were protected by state legislatures in teaching, languages from descendant's European origin (Crawford, 1992; Kloss, 1977; Menken, 2008). The fourteenth amendment to the United States' Constitution was ratified in 1868. This post-Civil War Reconstruction amendment states in part, "No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws" (United States Constitution Amendment XIV). Since its passage one hundred and fifty years ago, this amendment has served as the constitutional basis for the educational rights of language minority students (Wright, 2010).

The turn of the twentieth century brought new patterns of immigration, as well as a push back response of "Americanization" (Crawford, 1999; Menken, 2008). This rhetoric, accompanied by attempts at political action, emphasized an English only approach to public schools and institutions. This created particularly unique challenges for Native Americans that still persist to this day. "There are many students who have been exposed to another language and whose variety of English is colored by that language. Prominent among these are American Indian students, many of whom do not speak their heritage languages but whose English is nevertheless influenced by them (McCarthy, 2009)" (Basterra, Solano-Flores, Trumbull, 2011, p. 278). Native American experiences, a population with a significant number of ELs, have paralleled the history outlined with respect to the impact of language and education policy. Legislation over time has shifted between policies that revitalize native languages and those that

emphasize English-only accountability (Cohen & Allen, 2012). Similar to other EL populations it is an intricate balance with more than language at stake; culture and the interests of entire communities are impacted by the implementation of these policies.

In 1923, the Supreme Court case of *Meyers v. Nebraska* struck down these restrictive language laws, and made clear the Fourteenth Amendment's stature in providing protection for language minorities (Menken, 2008; Ricento, 1995). This era of education for ELs, lasting until the mid-1960s is most commonly identified as "sink or swim," and is reflective of a language as problem orientation towards bilingual education (Gándara, 2015; Menken, 2008, 2009; Ruiz, 1984). This perspective of language derives from the influential framework established by Ruiz (1984), proposing three basic orientations about language: language as a problem, language as a right, and language as a resource. Ruiz (1984) argues that governments, groups, and individuals all vary, both consciously and subconsciously in assumptions, regarding their orientation about language. He contends that many of the challenges concerning bilingual education arise from the tension inherent in the language as a problem and language as a right orientation.

The mid-1960s was a significant turning point for the federal government's role within education, and more specifically in the education of ELs. The 1964 passage of the *Civil Rights Act* was the first time that federal legislation recognized the unique language needs of ELs (limited English-speaking or limited English proficient students at the time) can lead to inequities in educational opportunity. Title VI of the Act prohibits denial of equal access to education because of a language minority student's limited proficiency in English (Menken, 2008; Gándara, 2015). In 1965, President Lyndon Johnson passed the *Elementary and Secondary Education Act* (ESEA). The ESEA is the primary federal law for funding K-12

education in the United States. Historically, the ESEA has been central in shifting a national focus to supporting the needs of ELs (Hopkins, Thompson, Linquati, Hakuta, & August 2013). The law was amended in 1968 to add Title VII, known as the *Bilingual Education Act* (BEA). The BEA sought to address the challenges of students not speaking, or speaking well, English, and provided funding to meet the needs of these students in bilingual and ESL programs (Gándara, 2015; Menken, 2008). These sets of laws shifted the role of the federal government in education and the national tone towards one of greater linguistic tolerance, though not without some unintended consequences (Gándara, 2015; Hopkins, Thompson, Linquati, Hakuta, & August, 2013). Gándara (2015) outlines challenges of the BEA to include: a constantly shifting definition of goals and purposes with respect to bilingual programming; a deficit versus asset-based approach to ELs; and conflicts between desegregation and bilingual education. Moran (1988) also notes that vagueness of purpose in the BEA led to a lack of understanding the intentionality of bilingual program design and the linguistic philosophies underpinning it. Despite these unintended consequences, the BEA is throughout the literature, heralded as a seminal piece of legislation for recognizing and beginning to address the challenges faced by ELs (Gándara, 2015; Hopkins, Thompson, Linquati, Hakuta, & August 2013; Menken, 2008).

With an equally important historical impact, is the 1974 Supreme Court case *Lau v. Nichols* (Gándara, 2015; Menken, 2008, 2009; Wiley & Wright, 2004). The case centered on the education of 1,856 Chinese-speaking children in San Francisco, who claimed they were being denied equal access to education because no accommodations were made to support their understanding of classroom instruction in English. The decision stated that school districts needed to take "affirmative steps" to provide educational access and address the challenges of

English learners (Gándara, 2015; Menken, 2008, 2009). Shortly after the *Lau* decision, Congress passed the *Equal Educational Opportunities Act* (EEOA) that further clarified the ruling, and required school districts to "take appropriate action" in addressing language barriers that impede equal educational access. The *Lau Remedies* followed the next year, 1975, and provided guidelines from the Office of Civil Rights of the US Department of Education, linking the inclusion of linguistically diverse students to the delivery of a "free and appropriate education" (Menken, 2008; Wiley & Wright, 2004). The landmark *Lau* decision is credited with the enforcement of and actual implementation of programming supporting the initial intentions of the BEA (Menken, 2008). The BEA was reauthorized in 1978 and again in 1984. Each reauthorization brought with it a redefinition of the rights for ELs (Gándara, 2015). However it wasn't until the reauthorization of ESEA in 1994 that a shift back towards greater restriction of linguistic diversity began to take hold in federal policy (Gándara, 2015; Menken, 2008).

The 1994 reauthorization of ESEA was entitled *Improving America's Schools Act*. While the law was favorable towards bilingual education, it represented a movement away from access and equity, and a significantly increased emphasis upon standards and accountability (Fowler, 2000; Kaestle, 2001; Menken, 2008; Riddle, 1999). The law also allowed for a greater inclusion of ELs under Title I, targeting support for school districts with high levels of poverty (Anstrom, 1995; Menken, 2008). While offering its support for native language instruction, the shift towards a framework of accountability opened the door for increased federal involvement in education (Menken, 2008, 2009).

No Child Left Behind (NCLB) was the 2002 reauthorization of the ESEA. The law's emphasis on accountability and high-stakes assessment was highly controversial (Abedi, 2003;

Hopkins, Thompson, Linquanti, Hakuta, & August, 2013; Menken, 2009). Researchers highlighted the positive intention of the law in disaggregating ELs from other student populations, to foster greater inclusion of ELs and other subgroups, as well as, bring attention to their language and academic needs (Hopkins, Thompson, Linquanti, Hakuta, & August, 2013), while also noting the unintended consequences and problematic reality of unrealistic proficiency targets and a framework of high-stakes accountability (Ho, 2008; Kortez, 2008; Linn, 2005, 2008). Other researchers viewed NCLB as significant federal overreach into the more localized domain of education, and an attempt to move towards an English-only approach to language acquisition (Crawford; 2002; Menken, 2008, 2009; Wiley & Wright, 2004). Despite varying interpretations of intentions across the research, consistent was the law's emphasis on standardized assessment within a system of high-stakes accountability. Additionally, NCLB mandated that ELs must take both academic content area and language proficiency assessments (Menken, 2008). Additionally, there was a clear shift away from the notion of "bilingual" in the law, as Title VII, the BEA, was removed from the legislation (Crawford, 2002; 2004; Evans & Hornberger, 2005; Menken, 2008; Wiley & Wright, 2004). Title VII was replaced by Title III of NCLB, also known as the *English Language Acquisition, Language Enhancement, and Academic Achievement Act,* removing any formal language mentioning "bilingual." Again having actualized the removal of the term "bilingual", the Office of Bilingual Education and Minority Language Affairs of the United States Department of Education was renamed the Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students (González, 2002). Consistent across the literature was whether purposefully structured or unintentional in consequence, NCLB has had a profound

impact on the educational and assessment experiences of ELs (Crawford, 2002; Ho, 2008; Kortez, 2008; Linn, 2005, 2008; Menken, 2008, 2009; Wiley & Wright, 2004).

The *Every Student Succeeds Act* (ESSA) was passed in 2015 as a replacement to NCLB, and is the most recent reauthorization to the 1985 Elementary and Secondary Education Act. ESSA is a shift from NCLB in that it has provided greater autonomy for states to determine accountability plans and goals. Though similar to NCLB, ESSA retains mandatory standardized testing (Whetzel, 2017). The new law is an evolution of the accountability framework in that it provides greater latitude to states in determining the indicators, goals, targets, rewards, and consequences for schools and districts within the system (Lyons & Dadey, 2017).

With respect to ELs, ESSA is consistent with NCLB in that states must have adopted English language proficiency standards aligned to academic standards, as well as, administer assessments for both English language and academic content area proficiency (Menken, 2008; Lyons & Dadey, 2017). However, ESSA also recognized the unique needs of ELs, particularly with respect to distinct groups of ELs such as English learners with disabilities, recently arrived ELs, and long-term ELs. The law shifted several provisions relevant to ELs from Title III to Title I of the ESEA, addressing Title III State formula grants, as well as, limited portions of Title I pertaining to ELs. These changes encouraged closer collaboration regarding the administration of Title I and Title III programming within schools and districts, as well as, provided the opportunity for increased protections and services to ELs. With respect to implementation, significant authority rests with each state in determining the interpretations of indicators and measures of growth, defining what it means to "progress in achieving English language proficiency" (Lyons & Dadey, 2017). In the State of Illinois, "English Learner Progress" is

weighted as five percent of a school's overall rating, and defined as the "percent of English Learners on track to attain language proficiency within five years of identification" (ISBE, 2018). The "on track to attain language proficiency" is determined by a student's annual performance on the ACCESS 2.0 for ELLs WIDA assessment, taking that score and benchmarking it against the expected average growth a student would need to reach a composite score of 4.8 after 5 years, and thus no longer qualify as an English learner in the State of Illinois.

As 2017-18 is the first school year that ESSA will be implemented throughout the country, there is limited current research as to its impact on teaching, learning, programming, and assessment. Whetzel (2017), Lyons and Dadey (2017) and others have provided overviews of the legislation with an emphasis on ELs, offering recommendations and perspective on how states can adopt policies that align with research findings.

Overall, this moment represents a timely opportunity for research regarding the impact of this legislation for ELs and the schools, districts, and communities that serve them. This review will now shift to explore the research concerning the stages, processes, and duration of second language acquisition for school age children.

Theoretical Considerations: Frameworks of Second Language Acquisition

This section of the literature review examines the learning theories that underlie teaching and second language acquisition within the United States. The following analysis of research for second language acquisition is intended to create a theoretical framework supporting the understanding of time, resources, support, cognitive ability, and environmental considerations necessary for a child to become proficient in English as a second language in the context of school. The construct of this theoretical framework serves to better understand the context of

second language acquisition for school aged ELs, as well as, the appropriateness of standardized assessment as a measure of content area understanding.

Cummins: Quadrants Framework and the Common Underlying Proficiency Model

In a progress report from the State of California Office of Bilingual Bicultural Education to educators regarding the programs and practices of bilingual students, Cummins (1981) synthesized research concerning the role of primary language development in the acquisition of English (Rivera, 1984). The author countered the common and widespread misunderstanding that a lack of English language proficiency is responsible for the academic failure of language minority students utilizing research studies, concluding the importance of primary language (L1) in secondary language (L2) development of proficiency (see Bereiter & Scardamalia, 1981; Donaldson, 1978; Olson, 1977). Cummins (1981) presents a theoretical framework for communicative proficiency relevant to bilingual students in the United States. These findings shifted the landscape of second language acquisition research and provided a foundation for other researchers in the field to build upon (Rivera, 1984).

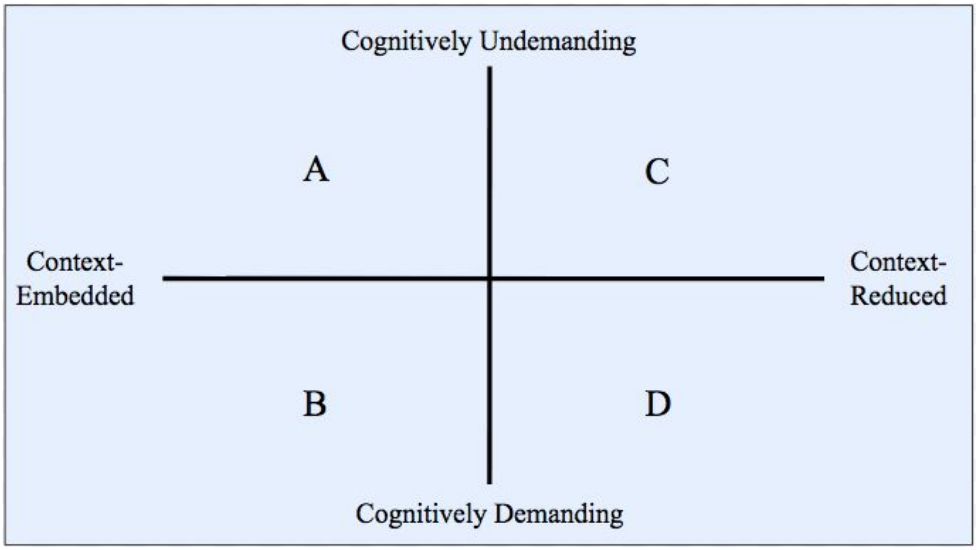


Figure 2. Cummins' Framework for Language Proficiency.

As displayed in *Figure 2*, this framework is centered on an x- and y- axis representing the continuums of contextual support and communicative competencies (Cummins, 1981). The intersection of these two axes creates a quadrant plane for situating communicative activities on the basis of their contextual and cognitive demand. Context-embedded communication derives from interpersonal involvement in a shared reality and reduces the need for linguistic elaboration of meaning, whereas, context-reduced communication does not assume this shared reality, and linguistic messages must be explicitly elaborated on to reduce misinterpretation (Cummins, 1981). The vertical axis places communicative activities along a continuum of cognitive demand. Communicative tasks that are more automatized, and thus mastered, appear higher on the axis, whereas more cognitively demanding tasks are situated below the horizontal axis. This quadrant plane provides a framework for identifying the cognitive demand and context embeddedness of communicative tasks, locating the level of challenge for such activities. Within the school environment, visual cues, linkage to a student's own experience and real life, and hands-on activities are all associated with greater context embodiment. Abstract language, oral language lacking support, and standardized assessment are all examples of context reduced activities within the classroom (Cummins, 1981,1989, 2000).

> What is not realized by many educators is that because of language minority students' ESL background, the regular English curriculum and psychological assessment procedures are considerably more context-reduced and cognitively demanding than they are for English-background students. (Cummins, 1981, pp. 15-16)

Furthermore, Cummins (1981) posits the Common Underlying Proficiency (CUP) theory based on findings that align with the research on second language development, explaining the

interdependence of literacy-related domains for the L1 and L2 language proficiency of bilingual students. According to the CUP model more cognitively demanding proficiencies (literacy, content learning, problem-solving) are common across languages. This theory provides the basis for Cummins' (1981) Linguistic Interdependence Hypothesis that all languages contain surface features, yet underlying these outward manifestations are proficiencies common to all languages. The understanding that L1 proficiency transfers towards the acquisition of L2 (Cummins, 1981). Cummins' CUP model is grounded in research exploring the effectiveness of bilingual education programming (Layba, 1978; Rosier & Farella, 1976) and empirical relationship between L1 and L2 proficiency (Cummins, 1981; Wells, 1979).

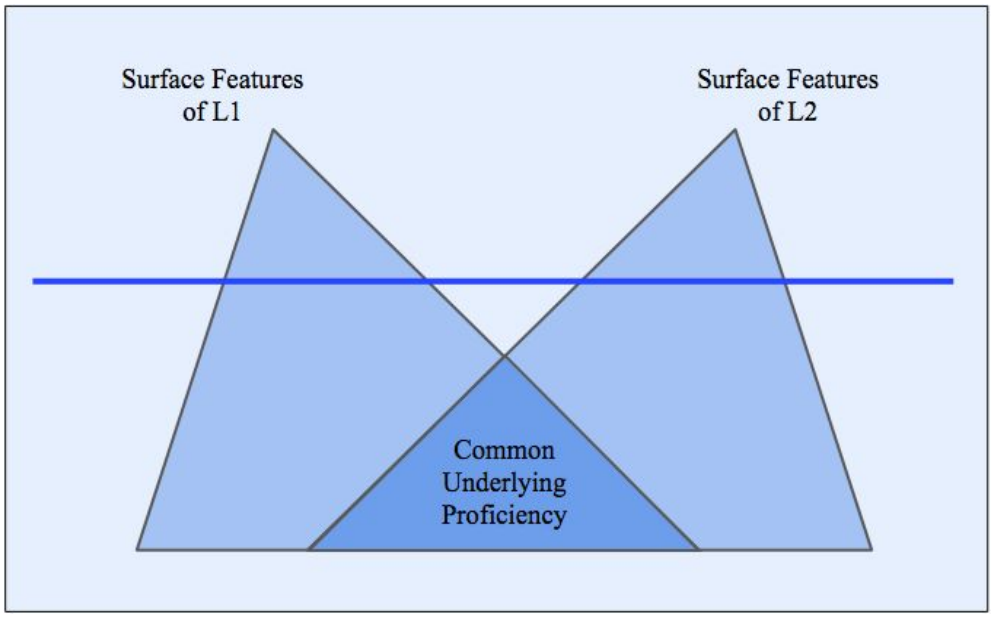


Figure 3. Cummins' Dual-Iceberg Model of Proficiency.

Figure 3 illustrates both the CUP model and Linguistic Interdependence Hypothesis. This visual representation models bilingual proficiency as a "dual iceberg" in which the commonalities of language acquisition for both L1 and L2 underlie the obvious surface differences between the two languages (Cummins, 1981). The iceberg visualization relates to the aforementioned

quadrants framework as surface level communicative tasks are less cognitively demanding and more context-embedded than those more challenging tasks lying below.

The language quadrants and CUP hypothesis created by Cummins (1981) represent a significant shift in the research on second language acquisition. This work emphasized an asset-based approach to developing second language proficiency, and encouraged educational systems to build upon the linguistic and cultural characteristics of students, rather than seeking to replace them. This research encourages teachers, schools, and educational systems to recognize the differences between context-reduced and context-embedded communication proficiencies and rethink their practices, programs, and policies accordingly (Rivera, 1984).

Collier & Thomas: The Prism Model

The Prism Model developed by Collier and Thomas is grounded in educational, linguistic, and social science research (Collier 1995a, 1995b, 1995c; Collier & Thomas, 2007; Ovando, Combs, & Collier, 2006). The model builds upon the aforementioned theoretical underpinnings of Cummins (1980, 1981) to offer a robust theoretical framework for second language acquisition. The Prism Model framework will provide a more complete theoretical context for how language acquisition plays out in schools and related educational policies.

The origins were developed by Collier (1995), and sought to clarify common misunderstandings between policymakers and educators in the United States with respect to the process of second language acquisition for students. Collier puts forth a prism model of language acquisition for school comprised of four major, and overlapping, components: sociocultural, linguistic, academic, and cognitive processes. The prism model builds upon previous research evidence, primarily that language acquisition is a complex, timely, and

unending process that occurs throughout one's life (Berko Gleason, 1993; Collier, 1992). Building upon Cummins (1981, 1989) and Krashen and Terrell (1983) research, Collier (1995a) reinforces the span of five to seven years for immigrant students to reach typical native-speaker performance, given 2-3 years of schooling in their first language. The amount of schooling in a child's first language is the most significant student background variable in determining progress towards second language proficiency. Cognitive and academic development in a child's first language is important and positive for second language development (Bialystok,1991; Collier, 1989, 1992; García,1994; Genesee, 1994; Thomas & Collier, 1995).

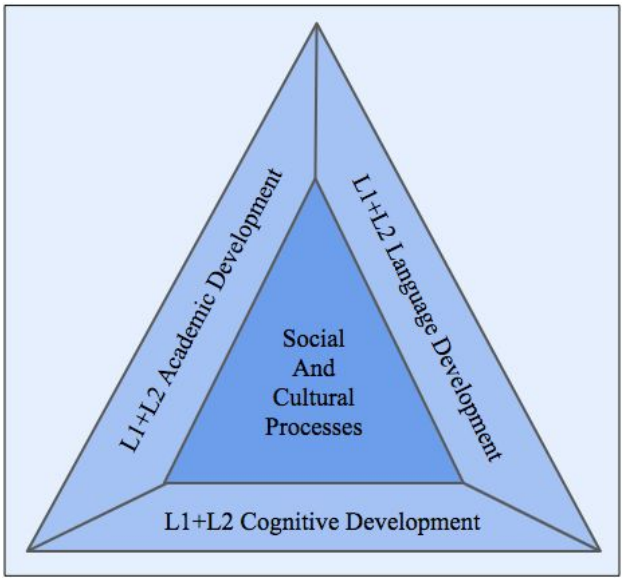


Figure 4. Collier & Thomas' Prism Model.

Collier and Thomas (2009) identify four primary components in the Prism Model that include: sociocultural processes, linguistic processes/language development, academic, and cognitive processes. As illustrated in *Figure 4*, academic, language, and cognitive development make up the exterior sides of the prism, while social and cultural processes lie at its core. Though represented two dimensionally, the model should be envisioned as a complex multi-dimensional

prism with the student at the center. A fundamental principle of the Prism Model is that linguistically diverse children, who are still developing their English language proficiency, require a school context that provides equivalent opportunities and advantages to the majority population (Collier & Thomas, 2009). The inclusion of sociocultural factors within a widely utilized model of second language acquisition represented a significant step forward from the aforementioned research. Sociocultural processes refer to the broader social surrounding and cultural context in which a student's life is situated, from home and school to the broader society. Linguistic processes are made up of the unconscious aspects of language development, inherent to all humans. Academic processes refers to work in the school environment, and the increasing demands both academically and linguistically with each succeeding grade. Cognitive processes are the natural, subconscious, development processes that occur from birth through schooling and beyond (Collier & Thomas, 2009). Similar to Cummins (1981) findings there exists interdependence between L1 and L2, as well as, between all four of these components.

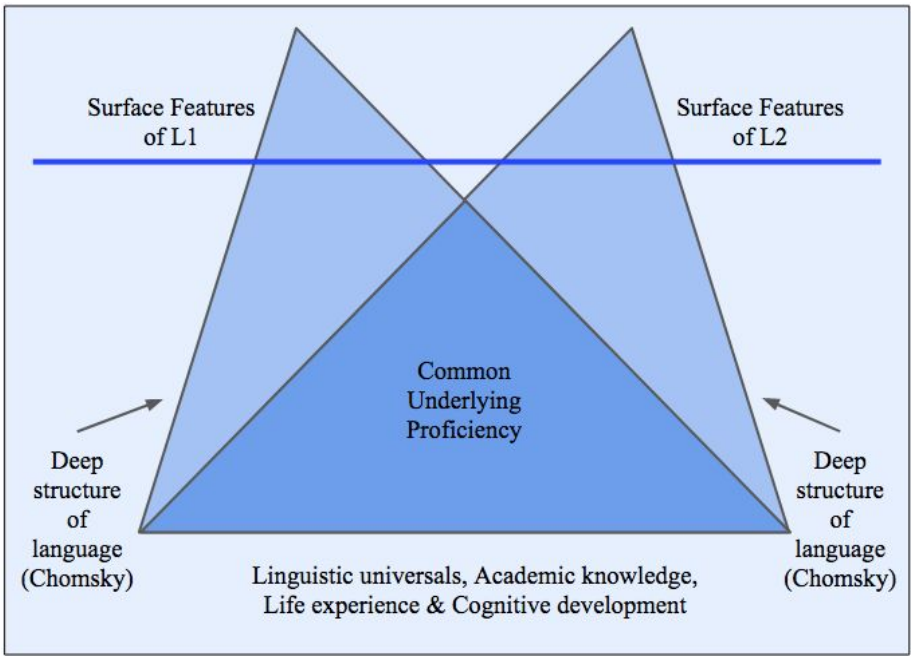


Figure 5. Cummins' Common Underlying Proficiency of L1 and L2 Pictorial "Dual-iceberg"

Figure 5 illustrates this interdependence between a student's first and second language, advanced by Cummins (1981) that deep knowledge transfers from the first language to the second. The iceberg model serves to illustrate that despite distinct surface level differences between two languages, the underlying deeper commonalities exist and serve to assist the development of a second language. Additional research points to many aspects in literacy that do not need to be retaught in a second language when students are literate in their first (Chu, 1981; Cummins, 1991, 2000; Goldenberg, 2008). Furthermore, the research also shows that cognitive processes developed in the first language also transfer to the second (Baker, 2006; Bialystok, 2001; González, 2005).

Based upon the Prism Model, Collier and Thomas (2009) recommend a bicultural context for schooling - integrating first language, culture, and community knowledge into the school day, and the importance of parents maintaining home language cultural traditions. Based on the research evidence, the authors recommend two-way bilingual as an effective program model. Additionally, for school environments that do not allow for that programming option, Collier and Thomas recommend classroom instruction grounded in problem-solving, discovery, and active engagement for students, as well as, that the second language be taught through academic content. Additional recommendations based on this body of research include ongoing professional development and training for staff with regard to the many overlapping processes and factors for second language acquisition (Collier & Thomas, 2009).

Summary of Theoretical Frameworks

This review of learning theories that underlie the approach to teaching and second language acquisition within the United States system provides a framework for understanding the

time, resources, support, cognitive ability, and environmental conditions necessary for a student to become proficient in English as a second language. This body of research contributes to the theoretical underpinnings from which the remainder of the literature will be reviewed. This theoretical framework provides a clear context of second language acquisition for school aged children. With a common understanding of the theoretical framework in place the review of literature will shift to focus on the research of how long it takes for ELs to develop native like proficiency.

Stages and Processes for the Second Language Acquisition of School Age Children

The length of time for students to acquire mastery of a second language has been one of the most important topics of study for the education of ELs over the past four decades (Cummins, 1981; Collier & Thomas, 2009; Hamayan & Freeman, 2012). The following studies have contributed significantly to the research findings on this matter. The "how long?" question builds upon the aforementioned theoretical framework in that it is based on how children learn a second language. This research is grounded in the manner in which children develop literacy in a second language, and how they develop cognitively and with respect to academic content through this second language (Hamayan & Freeman, 2012).

Understanding the "How Long?" Question

Throughout the literature this question is situated within a broader educational and socio-political context. Within the federal legislation of NCLB, English proficiency "refers to the ability to (1) meet the proficient level of achievement on state assessments; (2) successfully achieve in classrooms where the language of instruction is English; and (3) participate fully in society" (Cook, 2012). Across the literature, English proficiency encompasses not just the oral

proficiency of students, but a child's reading, writing, speaking, and listening proficiencies as measured by grade level standardized assessments across various content areas (Collier & Thomas, 1995, 2007, 2009; Cook, 2012; Cummins, 1981; Hakuta, Butler, Witt, 2000). In searching for an empirical answer to the "how long?" question, researchers have been largely driven by the length of time that it takes for ELs to reach typical performance, 50th percentile, of native English speakers within their age range on mandated standardized, norm- and criterion-referenced tests or performance assessments (Collier & Thomas, 2009; Cummins, 1981; Hakuta, Butler, Witt, 2000). For a significant body of the research "English proficiency" is defined by attaining average levels of performance on content-area assessments as compared to monolingual English speaking peers (Collier & Thomas, 2009; Cummins, 1981; Hakuta, Butler, Witt, 2000). *Figure 6* illustrates this concept and provides a visual representation of the time required ELs to close the gap with their English speaking peers.

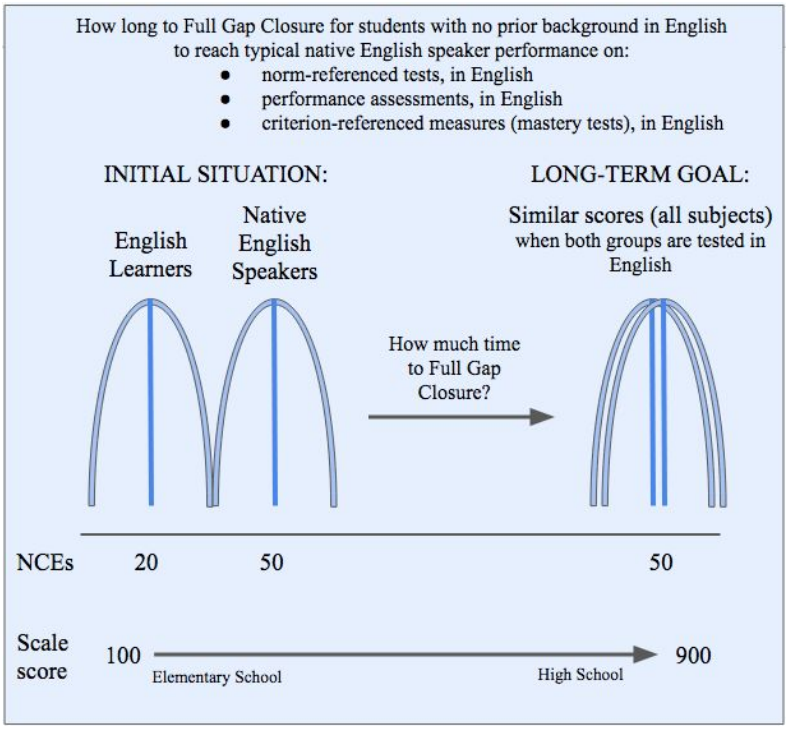


Figure 6. How Long to Close the Gap? Adapted from Collier & Thomas (2009).

Across the literature, the amount of time that it takes for ELs to acquire academic language proficiency is between five and seven years (Collier & Thomas 1989, 2005, 2009; Cummins, 1981, 2012; Hakuta, Butler, Witt, 2000). Cummins (1981) offers a groundbreaking starting point for exploring the research of second language acquisition, with respect to length of time to develop proficiency. Cummins (1981) study analyzed 1,210 immigrants in Canada, who at age six or younger were first exposed to English. In this longitudinal study, Cummins analyzed the data by age upon arrival and length of residence in Canada. In doing so he concluded that it took five to seven years, on average, for these students to approach grade-level norms on school assessments that measured cognitive-academic language development in English (Cummins, 1981). Influenced by the research of Cummins (1981), Collier and Thomas have since built upon this knowledge base, conducting studies in the United States regarding the length of time required for ELs to obtain native-like proficiency in English. They conducted studies in 35 school districts and 16 states, analyzing over 6.2 million individual student records. In synthesizing decades worth of research, Collier and Thomas (2009) conclude that five years is the shortest amount of time for the typical English learner to reach the achievement levels of an English speaker, and this given at least four years of on-grade level schooling in their primary language. On the other hand, students who receive English only instruction, never having received home language support or primary language schooling, may struggle to reach the same levels of achievement as their English speaking peers (Collier & Thomas, 2009). Evident throughout the research is the conclusion that the process of second language acquisition takes time, especially the development of academic language. At least five to seven years' time to

achieve native like English proficiency was a consistent finding across the literature (Cummins, 1981; Collier & Thomas 1989, 2005, 2009; Hakuta, Butler, Witt, 2000).

Cummins' (2012) findings show that each dimension of language (conversational fluency, discrete language skills, and academic language proficiency) requires a different amount of time to reach proficiency, and varies for each child. For ELs to become reasonably fluent in conversational English takes approximately one to two years. These findings concur with Collier (1987), which determined that ELs with multiple years of formal education in their primary language, still require at least two years to close the gap when compared with their monolingual English speaking peers. Cook, Boals, Wilmes, and Santos (2008) explain the evidence of this basic principle as "lower is faster, higher is slower" with regard to growth in language proficiency. Students beginning at lower proficiency growth levels will on the whole progress faster than students beginning at higher proficiency levels; lower grade students will generally grow faster than students in upper grades (Cook, 2012). Conversational fluency however is not the benchmark by which ELs are measured and held accountable to their native monolingual peers. Academic English is the language of school and also the language in which academic competence is assessed. Hence, Academic English is the language of achievement and thus ultimately success (Collier & Thomas, 2009; Cook, 2012; Cummins, 2012; Hakuta, Butler, Witt, 2000). Cummins (2012) further details the increased complexity of academic language with increasing grade level expectations, this concurs with studies conducted in various countries outside the United States. These findings are used as evidence to further support that second language learners typically require at least five years to reach the levels of native English speakers in academic English (Cummins, 2012).

Also, evident throughout the literature is the understanding that ELs must catch a moving target with the ongoing linguistic and academic growth of their English speaking peers (Cummins, 1981, 2012; Collier & Thomas 1987, 1989, 2007, 2009). To close the gap, ELs must not only match, but exceed the learning of their English speaking peers each and every year. See *Figure 7*, on average this requires ELs to make fifteen months growth in a ten month school year, every year, to close the gap in a five year period (Collier & Thomas 2009; Cummins 2012).

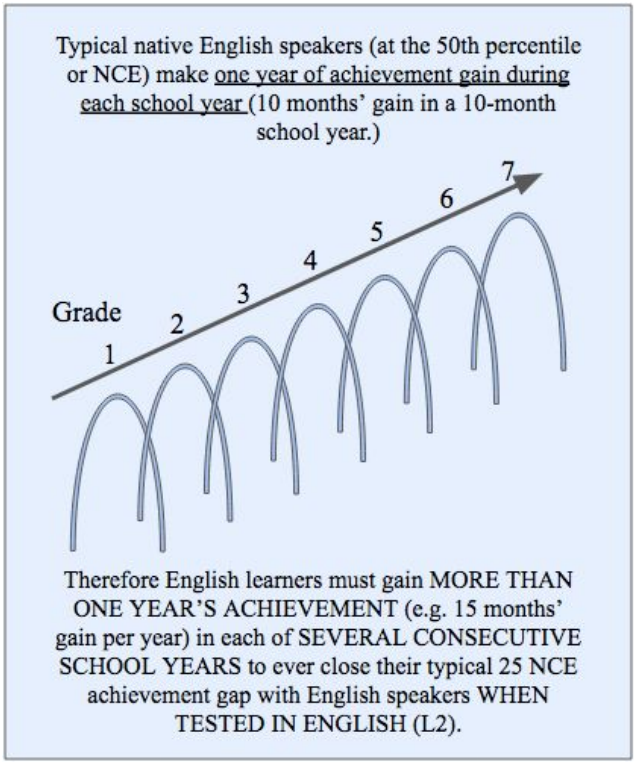


Figure 7. Closing the Gap. Adapted from Collier & Thomas (2009).

This widening gap between ELs and native language peers, highlights the immense task of developing both oral and academic proficiency, while keeping pace with monolingual classmates, who experience instruction in their primary language (Cummins, 1981; Collier & Thomas 1987, 1989, 2007, 2009; Hakuta, Butler, Witt, 2000). The literature emphasized many variables - age, length of residence, L1 proficiency, minority status, time in the school system,

type and quality of schooling, etc. - that influence the process of second language development and proficiency required to achieve academic success in the second language (Cummins, 1981, 2012; Collier & Thomas 1987, 1989, 2007, 2009; Cook, Boals, Wilmes, & Santos, 2008; Hakuta, Butler, Witt, 2000). A common research finding is that educators have to be mindful of where ELs start in their development of proficiency, as this greatly influences the length of time it takes for them to develop academic English proficiency and bridge the linguistic and academic gap with their English speaking classmates.

Stages of Second Language Acquisition

Building upon the aforementioned research that the amount of time required for a child to reach native like English proficiency varies depending on each child and various other factors, Krashen and Terrell outline a framework for the stages of this learning. According to Krashen and Terrell (1983) students acquiring a second language go through five stages of second language acquisition: preproduction, early production, speech emergence, intermediate fluency, and advanced fluency.

Stage	Characteristics	Approximate Time Frame	Teacher Prompts
Preproduction	The student: - has minimal comprehension - does not verbalize - nods "yes and "no" - draws and points	0 - 6 months	- show me... - circle the... - where is...? - who has...?
Early Production	The student: - has limited comprehension - produces one- or two-word responses - participates using key words and familiar phrases - uses present-tense verbs	6 months - 1 year	- yes/no questions - either/or questions - one- or two-word answers - lists -labels
Speech Emergence	The student: - has good comprehension - can produce simple sentences - makes grammar and pronounciation errors - frequently misunderstands jokes	1 - 3 years	- why...? - how...? - explain... - phrase or short-sentence answers
Intermediate Fluency	The student: - has excellent comprehension - makes few grammatical errors	3 - 5 years	- what would happen if...? - why do you think...?
Advanced Fluency	The student has a near-native level of speech.	5-7 years	- decide if... - retell...

Table 1. Stages of Second Language Acquisition. Adapted from Krashen & Terrell (1983).

Table 1 outlines these five stages, along with the key characteristics and approximate time frame for each stage. Note that the amount of time a child may spend at each stage will vary. Krashen & Terrell (1983) discuss the fluidity of second language acquisition and note that students will move between the various stages dependent on the cognitive demands of second language learning. This framework emerges from the theoretical model established by Cummins (1981), providing a more clearly defined process of second language acquisition for practitioners. The language of everyday communication is established first for students, typically within the first two years (Cummins, 1981; Collier, 1987, 1989; Krashen & Terrell, 1983). Returning to the Cummins' (1981) language quadrant framework and the iceberg analogy, highly context-embedded surface level language is at a low level of cognitive demand, and thus acquired in the earlier stages of development. Conversational ability is often misperceived by educators as a sign of fluency, while the deeper, more submerged levels of communication proficiency lie below the surface not fully developed. ELs will present as more fluent than they really are, requiring additional time, five to seven years to develop deeper, and more advanced, levels of fluency (Collier, 1987, 1989; Collier & Thomas, 2007, 2009; Cummins 1981, 2012; Krashen & Terrell, 1983). Escamilla (2012) outlines the linkage of this aforementioned research and stages of linguistic development with the present-day performance definitions and descriptors as part of the World-Class Instructional Design and Assessment (WIDA) framework, drawing a clear connection from the work Krashen & Terrell to the performance descriptors and ELD standards of today. The performance definitions outlined below describe the significant markers of language development, and represent a continuum of language learning rather than a leveled linear progression. *Table 2* describes receptive language and shows how ELs process

language to understand information, ideas, or concepts in oral or written communication. *Table 3* illustrates productive language and shows how ELs use language to express information, ideas, or concepts in oral and written communication (WIDA, 2012). These performance descriptors and the assessments that yield their data will be discussed further in subsequent sections of this review.

Figure D: WIDA Performance Definitions Listening and Reading, Grades K–12

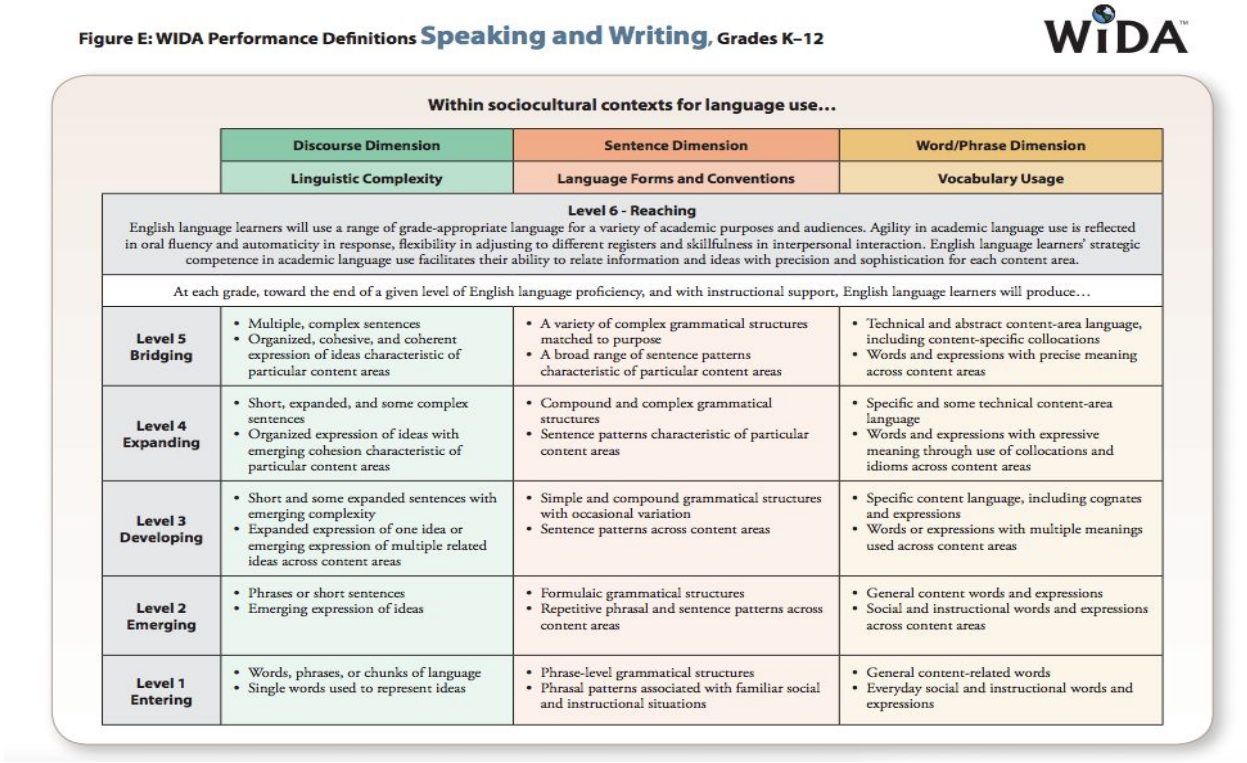
WIDA

Within sociocultural contexts for processing language…

	Discourse Dimension	Sentence Dimension	Word/Phrase Dimension
	Linguistic Complexity	**Language Forms and Conventions**	**Vocabulary Usage**
Level 6 - Reaching English language learners will process a range of grade-appropriate oral or written language for a variety of academic purposes and audiences. Automaticity in language processing is reflected in the ability to identify and act on significant information from a variety of genres and registers. English language learners' strategic competence in processing academic language facilitates their access to content area concepts and ideas.			
At each grade, toward the end of a given level of English language proficiency, and with instructional support, English language learners will process…			
Level 5 Bridging	• Rich descriptive discourse with complex sentences • Cohesive and organized, related ideas across content areas	• A variety of complex grammatical structures • Sentence patterns characteristic of particular content areas	• Technical and abstract content-area language • Words and expressions with shades of meaning across content areas
Level 4 Expanding	• Connected discourse with a variety of sentences • Expanded related ideas characteristic of particular content areas	• Complex grammatical structures • A broad range of sentence patterns characteristic of particular content areas	• Specific and some technical content-area language • Words or expressions with multiple meanings across content areas
Level 3 Developing	• Discourse with a series of extended sentences • Related ideas specific to particular content areas	• Compound and some complex grammatical constructions • Sentence patterns across content areas	• Specific content-area language and expressions • Words and expressions with common collocations and idioms across content areas
Level 2 Emerging	• Multiple related simple sentences • An idea with details	• Compound grammatical structures • Repetitive phrasal and sentence patterns across content areas	• General content words and expressions, including cognates • Social and instructional words and expressions across content areas
Level 1 Entering	• Single statements or questions • An idea within words, phrases, or chunks of language	• Simple grammatical constructions (e.g., commands, Wh- questions, declaratives) • Common social and instructional forms and patterns	• General content-related words • Everyday social, instructional and some content-related words and phrases

Table 2. WIDA Performance Definitions Listening & Reading, Grades K-12 (2012).

Figure E: WIDA Performance Definitions Speaking and Writing, Grades K–12



Within sociocultural contexts for language use…

	Discourse Dimension	Sentence Dimension	Word/Phrase Dimension
	Linguistic Complexity	**Language Forms and Conventions**	**Vocabulary Usage**
Level 6 - Reaching	English language learners will use a range of grade-appropriate language for a variety of academic purposes and audiences. Agility in academic language use is reflected in oral fluency and automaticity in response, flexibility in adjusting to different registers and skillfulness in interpersonal interaction. English language learners' strategic competence in academic language use facilitates their ability to relate information and ideas with precision and sophistication for each content area.		
At each grade, toward the end of a given level of English language proficiency, and with instructional support, English language learners will produce…			
Level 5 Bridging	• Multiple, complex sentences • Organized, cohesive, and coherent expression of ideas characteristic of particular content areas	• A variety of complex grammatical structures matched to purpose • A broad range of sentence patterns characteristic of particular content areas	• Technical and abstract content-area language, including content-specific collocations • Words and expressions with precise meaning across content areas
Level 4 Expanding	• Short, expanded, and some complex sentences • Organized expression of ideas with emerging cohesion characteristic of particular content areas	• Compound and complex grammatical structures • Sentence patterns characteristic of particular content areas	• Specific and some technical content-area language • Words and expressions with expressive meaning through use of collocations and idioms across content areas
Level 3 Developing	• Short and some expanded sentences with emerging complexity • Expanded expression of one idea or emerging expression of multiple related ideas across content areas	• Simple and compound grammatical structures with occasional variation • Sentence patterns across content areas	• Specific content language, including cognates and expressions • Words or expressions with multiple meanings used across content areas
Level 2 Emerging	• Phrases or short sentences • Emerging expression of ideas	• Formulaic grammatical structures • Repetitive phrasal and sentence patterns across content areas	• General content words and expressions • Social and instructional words and expressions across content areas
Level 1 Entering	• Words, phrases, or chunks of language • Single words used to represent ideas	• Phrase-level grammatical structures • Phrasal patterns associated with familiar social and instructional situations	• General content-related words • Everyday social and instructional words and expressions

Table 3. WIDA Performance Definitions Speaking & Writing, Grades K-12 (2012).

The performance definitions define expectations of student readiness for instruction at each level of English language proficiency. The definitions incorporate the criteria of linguistic complexity, vocabulary usage, and language forms and conventions for each level. It is with this research, and its more traditional cognitive-based linear approach to second language acquisition, that most schools in the United States ground their practices, pedagogies, and systems of support for ELs. This however, is not where research in the field ends, but rather where it really begins to take off.

The Expanding Divide Between School Practices and Current Second Language Acquisition Research

This section begins by situating current second language acquisition research within the broader historical context previously discussed. The emphasis of each of the following research

studies, and meta-analyses, provides a powerful opportunity to measure the field of second language acquisition at large and over an extended period of time.

In a review of second language acquisition research spanning the bulk of the twentieth century, 1916-2000, Chaudron (2001) identified key themes defining each significant era. Chaudron's analysis allows for a step back and an opportunity to better understand the broader scope of this research. Focusing on the last three decades of the twentieth century, provides a natural continuation of this section with those that preceded it, while highlighting prevailing themes of second language acquisition research over time. Accordingly, Chaudron (2001) identified the 1970s as a shift towards a focus on the individualization of instruction, the 1980s as seeking to understand cognitive processes of the learner, learners' psychology, and individual learning styles, and the 1990s as centered on the impact of language learning tasks, with an increased emphasis on sociocognitive perspectives.

Ellis (2012) informally extended this research into the first decade of the twenty-first century, noting the rise in interest of second language acquisition research, an expansion of sociocultural theories, and an increased focus on classroom interactions. Additionally, Ellis (2012) detailed the extensive expansion of the field and rapid output of publications during this time, as well as, the challenge this creates for synthesizing research given the immense scope.

In their survey spanning two decades, Stapleton and Shao (2018) also provide a broad overview of second language acquisition research that aligns with the historical themes and trends outlined (Chaudron, 2001; Ellis, 2012; Larsen-Freeman, 2018). Similarly, Stapleton and Shao (2018) emphasize the increased number of articles over the past five to ten years, and the combination of academic fields integrated into this research - pedagogy, psychology, and

linguistics, all infused with culture. This analysis also highlights the sociocultural turn in the research and language teaching as a whole, recognizing the classroom as a place for social interaction with concern for student thinking, beliefs, attitudes, and identity (Stapleton & Shao, 2018).

Situating current research within its historical perspective, Larsen-Freeman (2018) notes the division of research efforts into two distinct camps - with some researchers focusing on the basic process of second language acquisition, and others examining why different second language learners exhibit varying levels of success. These broader past trends have been documented and detailed in the aforementioned theoretical framework, stages, processes, and duration of second language acquisition. It is also important to note that discretion should be utilized with respect to some of the larger trends and themes identified.

These four extensive meta-analyses (Chaudron, 2001; Ellis, 2012; Larsen-Freeman, 2018; Stapleton & Shao, 2018) contribute consistent patterns in second language acquisition research and teaching over the past forty years. This section will now turn to the outgrowth of research within the past decade, emphasizing emergent ideas, and the chasm between theory and practice.

> Much language education is still based on a century-old model of the gradual acquisition of a new language through careful study over a number of years with the aim - for some - of reaching near native proficiency. Meanwhile, the reality is that people of all ages, and especially the mobile young, are managing to communicate across cultures and languages because they want to and need to, making use of prior knowledge, language, acquired online or through the media and electronic translation tools. (King, 2017, p. 34)

King's quotation elucidates the expanding divide between instructional practices at the classroom level and the widening field of second language acquisition research spanning the past decade. Baker & Wright (2017) describe this change as the education field, researchers, and educators moving beyond oversimplified and deficit-minded concepts towards more dynamic, fluid, and asset-based understandings of language development. Research in the last ten years provides varying explanations for this gap between theory and practice, ranging from broad and global scales to the local and school level. Beginning with a more expansive perspective, Kramsch (2014) describes globalization as changing the ways in which second languages are taught, learned, and utilized. This research describes the friction between a rapidly changing world and the monolithic institution of schools, "there has never been a greater tension between what is taught in the classroom and what students will need in the real world once they have left the classroom" (Kramsch, 2014, p. 296). Kramsch details the disconnect between the needed pedagogies and competencies of teachers today, with communicative language teaching, grounded in 1980s thinking, currently being utilized in schools. Similarly, Larsen-Freeman & Tedick (2016) describe this differential between research and practice, as a "one-size-fits-all approach" to language teaching in classrooms that serve students with diverse linguistic backgrounds and learning goals. Velasco & García (2014) discuss this gulf as schools being sites of little understanding for how multiple languages interact and impact learning. While critical, Baker & Wright (2017) also highlight the contributions of this initial research, Cummins in particular, for developing theories and ideas that helped to guide bilingual education programming, enact policy, and focus attention on the unique needs of ELs. As exemplified by this research (Baker & Wright, 2017; Kramsch, 2014; Larsen-Freeman, 2018; Larsen-Freeman &

Tedick, 2016; Velasco & García, 2014), situated across the literature is the overarching theme of school practices that are out of date and ill-equipped to respond to the multilingual needs of students in a rapidly changing global and local context.

Second language acquisition research of the last ten years is coherent on the many complex challenges that face students, teachers, and other actors within the educational systems as relates to language learning and application. Also prevalent throughout the literature is the proposal of pedagogies and practices for improving the alignment of language learning, school, and best-practice research. The theme of shifting the transmission of learning from a traditional teacher to student relationship, towards a more collaboratively developed communicative endeavor is consistent across the body of research (Baker & Wright, 2017; Canagarajah, 2011; García, 2013; Jessner et al, 2016; Kramsch, 2014; Larsen-Freeman 2018; Stapleton & Shao, 2018; Velasco & García, 2014). These ideas again range from the macro to micro level. Kramsch (2014) calls for pedagogy that is historically grounded, politically engaged, and reflective. Larsen-Freeman (2018) proposes a classroom that moves beyond merely reproducing native speaker competence to a more open learning environment, where students "acquire new voices and new ways of articulating experiences and ideas" (Larsen-Freeman, 2018, p. 64). Baker & Wright (2017) emphasize a more can do and asset-based educational experience, grounded in a new swell of bilingual brain research. Velasco and García (2014) envision this work as one in which students are provided differentiated pathways to process, create, and build academic content knowledge. A more human-centered and personalized approach is consistent across most of the second language acquisition research of the past decade.

Throughout the body of literature this human-centered emphasis is frequently referred to as an ecological orientation. An ecological orientation grounds the research in a more holistic and relational systems account, embracing the complexity of interconnected components in unexpected ways (Larsen-Freeman, 2018). This research perspective lends itself to a greater emphasis on the individual learner within a more personalized context and set of goals (de Bot, Lowie, & Verspoor, 2007; Larsen-Freeman, 2018). Kramsch (2014) builds upon this idea of increased individualization, while also noting the tension of what is taught in the classroom, and what is required once a student enters the real world. The ecological orientation is a part of the previously discussed thread in the research reflecting a greater emphasis on sociocognitive processes (e.g. García & Li, 2014; Kramsch, 2014; Kramsch & Zhang, 2018; Pugliese & Filice, 2012; Velasco & García, 2014).

This work is situated in a combination of cognitive involvement and social interaction theories (Larsen-Freeman, 2018). Throughout the research of the twenty-first century (e.g. De Bot et al.,2007; Larsen-Freeman & Cameron, 2008; Aronin & Jessner, 2015), the open-ended exploration of multilingual systems and the many factors that influence them, such as social and psychological ones, has come to be known as dynamic systems or complexity theory (DCT). DCT posits "contact between two or more language systems does not merely result in an overlap of systems, but causes a complete metamorphosis of all the (language) systems involved" (Herdina & Jessner, 2002, p. 116). Building upon this prior research Larsen-Freeman (2008, 2017) proposes an adaptive view of language learning called complex dynamic systems theory (CDST). CDST is a metatheory that views language as a dynamic system that is constantly changed through use. With respect to language acquisition, learners seek to make meaning by

adapting and innovating their linguistic capabilities (Larsen-Freeman, 2018). DCT, CDST, and other similar theories, are complex, sociocognitive, and nonlinear. These theories differ from historical cognitive-based second language acquisition research in that patterns are favored over rules as the focal point of linguistic learning.

Another common theme throughout the literature is the boundary between language use and acquisition becoming more blurred and personalized. For Larsen-Freeman this dynamic approach begins to address the "inert knowledge problem" of students' not being able to apply their learning outside of a particular lesson. Kramsch describes this as one being unable to study the process of learning apart from the learner. Building upon previous research (Gass, 1998), Kramsch (2014) and others (Larsen-Freeman, 2018; Velasco & García, 2014) detail the difference between language learning and language use. Bunch (2014) presents this dichotomy as the language of ideas and language of display, explaining how students make use of their linguistic tools when engaged in an academic task. Similarly and taking into consideration the classroom perspective, Baker & Wright (2017) note this as "the distinction between teaching a language and teaching through a language" (p. 274). This line of thinking differentiates itself from past research that frequently collapsed the two together, and has led to new ways of exploring the interrelationship between language practices.

Over the past decade, the interconnection of linguistic practices has continued to evolve and progress, spanning many forms and styles - flexible bilingualism (Blackledge & Creese, 2010), heteroglossia (Bailey, 2007), polylingualism (Jorgensen, 2010), metrolingualism (Otsuji & Pennycook, 2010), and code meshing (Michael-Luna & Canagarajah, 2007; Canagarajah, 2011) overview earlier iterations of this research. The evolution of this work has led to the

creation of translanguaging as a pedagogy to incorporate multilingual communicative strategies within the educational environment. Translanguaging has become popular, perhaps even trendy, within the language acquisition literature, though it has been slow to gain widespread use within most educational systems. The term was initially conceived by the Welsh educator, Cen Williams (1996) and developed further by García (2009) and other researchers (Blackledge & Creese, 2010; Canagarajah, 2011; Lewis, Jones, & Baker 2012).

Cummins' interdependence theory underlies translanguaging practices, and translanguaging extends beyond Cummins' research in that L1 and L2 are not simply static. The theory of translanguaging builds upon the aforementioned research of Cummins (2009) and what he refers to as "bilingual instructional strategies" to build new pedagogical practices for co-constructing communication (Velasco & García, 2014). In contrast to much of the earlier research within the field, translanguaging does not view the languages of bilingual students as separate linguistic systems. Rather, it emphasizes the flexibility and meaningful actions through which bilinguals select linguistic strategies to communicate appropriately (Velasco & García, 2014).

Translanguaging builds upon the previous notion of blurring the line between language use and application. Baker and Wright (2017) describe translanguaging as "the planned and systematic use of two languages inside the same lesson..the input (reading and/or listening) tends to be in one language, and the output (speaking and/ore writing) in the other language, and this is systematically varied" (p. 280). García (2009) establishes translanguaging as a framework for conceptualizing the education of bilinguals as a democratic aim for social justice. Similar to other socioculturally grounded and social-justice oriented research the aim is to create an

educational space "that legitimate the intelligence, imagination, and linguistic talents of ELL students" (Velasco & García, 2014, p.8). Canagarajah (2011) builds upon these ideas, noting that translanguaging is a naturally occurring reality for bilingual students, though is careful to note that there is still further language to be developed. Baker and Wright (2017) highlight four potential advantages of translanguaging and transliteracy: may promote a more comprehensive understanding of the subject matter, may help develop oral communication and literacy in L2, dual language use can strengthen the home-school partnership, and greater integration of English speakers and English learners within classroom instruction. As reflected above, there has already been a considerable foundation of research conducted on translanguaging within classrooms, particularly with respect to language blending and social uses (Baker, 2003; Baker & Wright, 2017; Canagarajah 2007; García, 2009; Gorter & Cenoz, 2010), there still exists a gap in the literature with respect to literacy, including an emphasis on writing, and language for discourse and persuasion (Canagarajah, 2011).

Through its complex, nonlinear, sociocognitive, personalized, human-centered, and social-justice oriented approach, translanguaging reflects many of the shifts within the field of second language research over the past decade. Translanguaging as a pedagogy for building the linguistic and communicative capabilities of students is also reflective of the aforementioned broader theme of the expanding divide between instructional practices at the classroom level and the widening field of second language acquisition research. This body of research has expanded in extensive and meaningful ways over the past ten years (Baker & Wright, 2017; Chaudron, 2001; Ellis, 2012; Larsen-Freeman, 2018; Stapleton & Shao, 2018). It is now incumbent on practitioners to begin applying its findings with the increasing number of students who need this

support. The following section of this review will address critical issues in the assessment of ELs.

Critical Issues in the Assessment of ELs

The research findings concerning the assessment of ELs highlight a substantial difference in the academic performance of these students from that of their native English speaking classmates (Abedi, 2004, 2006; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Solano-Flores & Trumbull, 2003). As was discussed in the previous section regarding the processes of second language acquisition, ELs face significant hurdles in learning both English and academic content in English at the same time. Emphasizing the key research finding that it takes five to seven years for most ELs to gain mastery of academic English at a level comparable to native English speaking peers (Collier & Thomas 1989, 2005, 2009; Cummins, 1981, 2012; Hakuta, Butler, Witt, 2000). The burden of complex language on assessments is a significant factor that contributes to the gap in performance between ELs and their non-EL peers (Abedi, 2006; Abedi, Hofstetter, & Lord, 2004; Solano-Flores & Trumbull, 2003). The overarching theme of EL performance on standardized assessments being significantly impacted by language proficiency and linguistic complexity rather than content knowledge is paramount across the research (Abedi, 2002, 2003, 2009, 2011; Solano-Flores, 2008; Trumbull & Solano-Flores, 2011).

The most universal factor impacting the assessment of ELs is that standardized assessments are not designed, nor implemented, with the EL student population in mind (Abedi, 2011; Kopriva, 2000; Solano-Flores, 2011; Trumbull et al, 2011). Validity and reliability are central to the effectiveness of any assessment. Abedi (2011) defines reliability as "the ability of

a test to produce consistent measures or the ability of scorers or scoring machines to come up with the same scores for each item" (p. 58). Validity concerns for ELs are two-pronged: "(1) the construct being measured and (2) the structural relationship of assessment outcomes across students' ELL language status" (Abedi, 2001, p. 58). Whether a test measures what it intends to, and does so in a consistent manner is at the heart of any meaningful assessment. Increased linguistic demand decreases the validity and reliability of assessments for ELs (Abedi, 2002, 2006; Solano-Flores, 2008). When these language demands are lessened the performance gaps between ELs and non-ELs begin to close (Abedi, 2009). Additional research further supports that language factors have a more significant impact on ELs than on native English speakers (Abedi, 2006; Abedi et al, 2003). For ELs the greatest source of measurement error, and thus reduced assessment validity, comes from unnecessary linguistic complexity (Abedi, 2002, 2006; Solano-Flores, 2008). Unnecessarily dense language serves to distort the construct of individual questions and thus the entirety of assessments for ELs, measuring their language proficiency rather than the intended content objectives (Abedi, 2011). Throughout the research these unrelated variables that impact the measurement outcomes of assessments are referred to as "nuisance variables" (Abedi, 2006), "contaminants" or "construct-irrelevant variables" (Haladyna & Downing, 2004; Messick, 1994), and "extraneous variables" (Linn & Gronlund, 1995). Irrespective of name, the research is consistent in its findings that these variables jeopardize the validity and reliability of assessments for ELs. Furthermore, a consistent finding among researchers is that language factors unrelated to the content construct of assessments are amongst the most significant "nuisance" variables impacting the performance of ELs (Abedi, 2011). The Standards for Educational and Psychological Testing (American Educational

Research Association [AERA], & National Council on Measurement in Education [NCME], 1999) reinforce this core principle:

> Test use with individuals who have not sufficiently acquired the language of the test may introduce construct irrelevant components to the testing process. In such instances, tests results may not reflect accurately the qualities and competences intended to be measured. [Therefore] special attention to issues related to language and culture may be needed when developing, administering, scoring, and interpreting test scores and making decisions based on test scores. (p. 91)

Linguistic demands are inherent in standardized assessments taken by ELs as there is a presumed level of English language proficiency in the construct of these tests that are not related to the content knowledge being assessed (Trumbull & Solano-Flores, 2011). Abedi (2003) linked the impact of this linguistic demand with the performance of ELs on standardized achievement tests and concluded that greater language usage expanded the performance gap between ELs and their English speaking peers. A multiple-group confirmatory factor analysis model by Abedi et al (2003) found that assessment outcomes for ELs are not structurally as consistent as those of non ELs due to linguistic demand as a source of construct irrelevance. Further research has identified specific sources of language demands - vocabulary, syntax, discourse - that create issues for ELs and other students by conflating knowledge and skills with language proficiency (Abedi, Hofstetter, Baker, & Lord, 2001; Trumbull & Solano-Flores, 2011). With respect to vocabulary, false cognates, unfamiliar words and phrases, and lengthy words all lead to increased linguistic demand. Syntactically, unfamiliar tenses, negation, prepositional phrases, and compound sentences are amongst several grammatical considerations that increase the linguistic

requirement of an assessment. Lengthy problem statements, multiple instructional steps, and passive voice all contribute to the construction of assessment discourse that add an extraneous language burden for ELs (Abedi, Hofstetter, Baker, & Lord, 2001; Trumbull & Solano-Flores, 2011). These linguistic demands, independent of other cultural demands, add to the complexity of assessment for ELs and stand as a barrier for gauging their true depth of knowledge and understanding. As Trumbull and Solano-Flores (2011) note, "it is impossible to examine language in any depth without consideration for the culture that has created it, nor is it possible to gain deep understanding of a culture without knowing something about its language" (p. 27). This quote underscores a central tenet of the research that culture and language are deeply embedded and intertwined in standardized assessment practices for ELs (Abedi, Hofstetter, Baker, & Lord, 2001; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Trumbull & Solano-Flores, 2011). Specific issues regarding the cultural and linguistic demands of testing for ELs will now be discussed.

The Cultural and Linguistic Validity of Assessment Practices for ELs

This section of the literature review situates standardized assessments within a socio-cultural context and further explore the validity of these practices. Similar to any other human construct, tests are cultural artifacts (Cole, 1999; Solano-Flores, 2011). Standardized assessments in particular are part of a broader and complex system of instructional practices. These practices are connected to societal norms and expectations of accountability and responsibility, intertwined within legislation and policy at the local, state, and federal level (Solano-Flores, 2011). As was discussed in previous sections, academic English is the language

of assessment; the language of those who create, write, and evaluate these tests. This language is then used to create content that reflects the knowledge and values of those in society with power (Solano-Flores, 2011). Exploring testing as a cultural artifact allows researchers the opportunity to explore thinking through the student perspective and better understand the cultural factors that impact the assessment experience. Pulling together these various social, cultural, and political contexts, Solano-Flores and Nelson-Barber (2001) propose the concept of cultural validity as an independent form of validity for assessments. Cultural validity is defined as:

> The effectiveness with which [...] assessment addresses the socio-cultural influences that shape student thinking and the ways in which students make sense of [...] items and respond to them. These socio-cultural influences include the set of values, beliefs, experiences, communication patterns, teaching and learning styles, and epistemologies inherent in the students' cultural backgrounds, and the socioeconomic conditions prevailing in their cultural groups. (p.555)

Cultural validity builds upon a significant body of earlier research focused on the examination of instructional and assessment practices from a cultural perspective (Ladson-Billings, 1995; Kirkhart, 1995; Roseberry, Warren, & Conant, 1992). Solano-Flores & Nelson-Barber (2011) view these cultural factors as so ingrained and complex to an individual's thinking that they are intrinsic to any assessment practice.

Language is the principal way in which humans communicate with and learn from one another; the system by which we construct new ideas, knowledge, and understandings (Bronkart, 1995; Mantero, 2002; Trumbull & Solano-Flores, 2011). Language is a cultural construct and regarded as such throughout the research. As was detailed in earlier sections of this review,

language is inextricably linked to classroom instruction and assessment practices (Abedi, Hofstetter, Baker, & Lord, 2001; Trumbull & Solano-Flores, 2011). Both foundational and more specialized linguistic skills are imperative to the academic success of children within classrooms and on the standardized assessments intended to measure their core content knowledge (Trumbull & Solano-Flores, 2011). Consistent throughout the fields of sociolinguistic and assessment research of the past two decades is the notion that for ELs testing is more a function of language than of academic content-learning, and may reflect very little of what a child is fully capable of (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Heubert & Hauser, 1999).

Another key theme across the research is that standardized assessments, designed for native English speakers, do not provide valid and reliable outcomes for ELs (Abedi, 2011; August & Hakuta, 1997; Durán, 2011; Garcia & Pearson, 1994). The construct and design processes of standardized assessments are not inclusive of language factors and do not fully consider the multifaceted linguistic needs of ELs (Abedi, 2011; Kopriva, 2000; Rivera & Vincent, 1997). While these assessments may be constructed with universally accepted design principles and may indeed be valid and reliable for native English speakers, they are not necessarily so for ELs (Abedi, 2011). Durán (2011) notes that assessments and the blueprints guiding their construction cannot fully represent a specific learning domain and can only provide partial evidence of the range of mastery that a student has relative to specific content objectives. Hence, a student's performance is determined not only by their knowledge, but also based upon their interpretation of the assessment task they are being asked to perform (Durán, 2011). Critical issues regarding the format and scoring of assessments for ELs will now be reviewed.

Assessment Formatting and Scoring Considerations

In further exploring the body of research concerning critical issues in the assessment of ELs, it is important to understand the types of questions and overall formatting used on large-scale assessments, as well as, the impact upon student performance and scoring. Research shows that the alignment, or lack thereof, of assessment questions and formatting with the everyday instructional practices of classrooms may be more significant for ELs than for their native English speaking peers (Durán, 2011; Menken, 2000; Trumbull & Koelsch, 2011).

Large-scale standardized assessments, like PARCC, typically feature a combination of on-demand selected and constructed response items. These questions are presented on assessments to students in a prearranged and standardized method, with a fixed order of presentation (Durán, 2011). Most standardized assessments are taken independent of other instructional activities and academic tasks within the classroom, with students generally asked to respond to a series of back-to-back questions that are sequentially separated from one another (Durán, 2011). The multiple-choice format of one "correct" answer and other "incorrect" answers is the most prevalent on-demand selected response item on large-scale assessments within the United States educational system (Durán, 2011; Menken, 2008). This format creates several issues for ELs. Multiple-choice testing is a common experience for students in the United States beginning at the earliest grades. Students with this early exposure and practice, either on their own or through explicit instruction, develop test taking strategies, e.g, removing nonsensical or blatantly incorrect answers, that improve their overall performance on such assessments. Students with educational backgrounds outside of the United States, many ELs, not only lack this assessment exposure and practice, but may have been asked to demonstrate their

knowledge in vastly different formats, thus creating an inherent disadvantage when it comes to a multiple-choice assessment (Durán, 2011). Single correct response multiple-choice items are also problematic for ELs as many "incorrect" answers are constructed as "distractors," answers within reason, but not necessarily the "best choice." This assessment dynamic is a particular challenge for ELs as other answers may seem reasonable, and thus selected as correct, given a developing academic vocabulary and cultural context. As was discussed previously regarding the linguistic complexity of assessments, many questions intend to assess content objectives irrespective of other vocabulary, yet for ELs their understanding of that vocabulary is less fully developed than that of their native English speaking peers (August, Carlo, Dressler, & Snow, 2005). Durán (2011) details the impact of this on-demand selected response item issue with the overall performance of ELs on standardized assessments:

> The wording of assessment items will vary in terms of their linguistic complexity and the cultural load of their content...ELL students who do not understand a problem item adequately may not respond to the item at all or respond by selecting an answer option randomly. Both of these outcomes can lower the overall score earned by an ELL student by adding error to the estimate of his or her proficiency in the domain under assessment. (p. 126)

ELs face similar challenges with on-demand constructed response items as well. These question types can vary in length from short to extended depending on the expectations for the amount of writing required in response to the prompt. The issue of ELs responding to the question as intended, even with mastery of the content and skills, is a significant factor limiting the performance, validity, and reliability of standardized assessments for ELs (Abedi, 2002, 2006;

Durán, 2011). ELs must have adequate linguistic and cultural knowledge to be able to navigate the assessment and understand the demands being placed upon them (Durán, 2011; Solano-Flores & Nelson-Barber, 2001; Solano-Flores, 2011).

Scoring is another critical issue facing ELs for on-demand selected and constructed response items. While on-demand selected response items are scored electronically, the overall presentation of such data is one of objectivity and neutrality - the student's answer is either "correct" or "incorrect". For the reasons discussed above, that does not accurately reflect the assessment experience of ELs in relation to their native English speaking peers (Durán, 2011). On-demand constructed response questions pose a different issue for ELs in that they are typically evaluated by a scorer relative to a pre-constructed rubric. This method creates issues unique to ELs as performance is impacted by the scorer's ability to interpret the written response in the manner intended by the student (Durán, 2011). In order to do so effectively for an EL, a scorer would need to be knowledgeable of the ways that a child's response would be impacted by their language and cultural background, separated from the content objectives being assessed (Durán, 2011; Kopriva & Sexton, 2011; Lara & Chia, 2011; Trumbull & Solano-Flores, 2011). Furthermore, each of these issues related to formatting and scoring vary amongst ELs on a student by student basis. Depending on background and prior educational experience some ELs may be more ready and adequately prepared to manage the rigorous demands of constructed-response items and standardized assessments. As Durán (2011) states, "it is based not only on students' familiarity with the skills and knowledge they are expected to exercise but also their linguistic, cultural, and educational backgrounds" (p. 130). The intersectionality of

language, culture, and life experience is central to the literature surrounding ELs and assessment practices.

For all students, ELs and non-ELs, it is essential to understand that a core attribute of an assessment score is that it is an estimate of what a student knows and can do in the domain being assessed for one particular moment in time (Durán, 2011; Koretz, 2008). Large-scale standardized assessments are scored in various ways, weighting subscores to varying degrees, that in turn provide greater or lesser significance to different domains depending on the assessment. Typically, scores are provided on a numerical scale and may also be presented as proficiency levels, bands, or ranges (Durán, 2011). Overall assessment scores are estimates of the underlying content knowledge and skills that a student has, and is impacted by both construct relevant variability and construct irrelevant variability, with language factors being the most significant determinant for ELs (Abedi, 2002, 2006; Durán, 2011; Solano-Flores, 2008). Throughout the research, the most prevalent recommendation for better understanding the standardized assessment scores of ELs is to thoroughly evaluate the information through multiple lenses within a context unique to each particular child, school, and community (Durán, 2011).

Accommodations Issues for ELs

In continuing to examine the research regarding critical assessment issues, it is important to consider the context in which ELs take these large-scale standardized tests, as well as, the appropriateness and effectiveness of such accommodations. Though the definition of test accommodation varies, their function is as a "support provided students for a given testing event either through modification of the test itself or through modification of the testing procedure to help students access the content in English and better demonstrate what they know" (Butler &

Stevens, 1997, p.5). According to a joint statement of the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, "if English language learners are tested in English, their performance should be interpreted in light of their language proficiency. Special accommodations for English language learners may be necessary to obtain valid scores" (AERA, APA, & NCME, 1999, p.5). Furthermore, according to the Standards for Educational Accountability Systems, the intention of testing accommodations is to eliminate areas of complication that are unconnected to what is intended to be measured (Abedi, Hofstetter, & Lord, 2004; Baker, Linn, Herman, & Koretz, 2002). The overall goal of such accommodations is to create a more equitable experiences for ELs, without granting them an advantage over other students who do not receive assessment accommodations (Abedi, Hofstetter, & Lord, 2004; Baker et al, 2002; Thurlow, Liu, Erickson, Spicuza, & El Sawaf, 1996). Nearly every state allows some form of test accommodation on large-scale standardized assessments. It is the appropriateness, validity, and impact of such accommodations for ELs that emerge as a consistent critique across the research (Abedi, 2007, 2011; Abedi, Hofstetter, & Lord, 2004; Lara & Chia, 2011).

The primary reason for the critique of EL assessment accommodations across the literature is concern over the effectiveness and appropriateness of specific accommodation types. This concern is grounded in that many of the accommodations being provided or recommended to ELs were initially developed for special education populations, and fail to meet the unique needs of the EL population (Abedi, Hofstetter, & Lord, 2004; Lara & Chia, 2011). As such, these accommodations may not be appropriate or particularly effective for ELs, and may even lead to assessment outcomes that are not valid (Abedi, 2007; 2011). The research also highlights

a growing discrepancy between the use of allowable accommodations that have been demonstrated to appropriately support ELs and those that are most frequently provided as part of standardized assessments. The most frequently utilized accommodations are those least recommended for ELs: extra time, small group administration, linguistic modification of testing prompts, access to English-only dictionaries or glossaries, and reading the test aloud (Abedi, Elon, & Mirocha, 2003; Butler & Stevens, 1997; Heubert & Hauser, 1999; Shepard, Taylor, & Betebenner, 1998; Trumbull & Koelsch, 2011). Translating test items from English into other languages is not an effective accommodation strategy when English is the language of instruction in the classroom. The language of assessment should match the primary language used for learning within the school setting (Abedi, Hofstetter, & Lord, 2004). Across settings, extra time is the most common testing accommodation (Abedi, Hofstetter, & Lord, 2004). Abedi, Lord, Hofstetter et al, (2000) found that ELs benefited from extra time, as did English proficiency students. Other research studies found that though frequently provided as an accommodation, extra time was not always effective (Hafner, 2001; Kopriva, 2000). Oftentimes these accommodations were only provided to students at the most basic levels of English proficiency (Abedi, Leon, & Mirocha, 2003). These accommodations can be classified as timing, scheduling, and setting accommodations, and fail to specifically address the linguistic needs of ELs (Menken, 2000). Accommodations less frequently provided, but deemed most appropriate by empirical research include: modifying assessment vocabulary, providing additional examples, and modifying the overall linguistic complexity of the assessment (Abedi, Hofstetter, & Lord, 2004; Lara & Chia, 2011). This disconnect between practice and research regarding assessment accommodations for ELs ties back to a key point of the previous section,

that ELs are not considered at the outset of creating these standardized assessments. As Trumbull et al (2011) state, "adaptations and accommodations are actions taken after an assessment has been developed to make it more appropriate for particular students. However, the best hope for cultural validity is to consider the diversity of the population to be assessed at the very beginning of the development process-whether for large-scale or classroom assessment" (p. 283). This notion drives the findings and recommendations of researchers, that ELs must be taken into consideration at the beginning of the assessment design process and not simply accommodated afterwards.

Current research efforts, led by Abedi and others, have expanded the general understanding of assessment accommodations most appropriate to ELs. In a review of empirical research concerning accommodation practices, Abedi (2004) strongly advocates against a "one size fits all" approach to EL testing accommodations. As has been detailed throughout this review, ELs are an incredibly diverse population, and the effectiveness and validity of any assessment accommodation depends on meeting the unique needs of each particular child. Consideration should be given to a range of variables - English language proficiency, time in the United States, reading proficiency, level of schooling in home language, as well as, other demographic factors that statistically impact standardized assessment scores, socioeconomic status, parent educational levels, etc. (Abedi et al, 2004). Additional recommendations from the research include that all students, ELs and non-ELs, benefit from clear assessment language. Language can be further modified for ELs by reducing the use of low-frequency vocabulary and complex language structures that are irrelevant to the content being tested. This accommodation is an effective and valid practice as it does not impact the performance of English proficient

students (Abedi, Hofstetter, Lord, 2004; Lara & Chia, 2011; Trumbull & Koelsch, 2011). Personalized dictionaries are identified as an effective support for ELs, that also do not affect the score of English proficient students. Again all students could benefit from such accommodation (Abedi, Hofstter, Lord, 2004). Addressing the accommodation related issues of assessment for ELs is one more piece to ensuring a more equitable assessment experience. As Trumbull & Koelsch (2011) note, "achieving validity and equity in the assessment of ELLs requires addressing all of the elements of assessment: content (including language of instructions and any other text), format, administration, scoring criteria and procedures, score interpretation, and use" (Farr & Trumbull, 1997; Solano-Flores et al, 2001; Solano-Flores & Trumbull, 2003; Solano-Flores, Trumbull, & Nelson-Barber, 2002, p. 196). It will be in addressing the totality of these factors, that the assessment process becomes a more valid, equitable, and ultimately meaningful experience for ELs and the school communities that serve them.

Conclusion

This chapter examined the theory and research concerning second language acquisition and assessment of ELs within the context of public education in the United States. Furthermore, this review of the literature summarized the practices and policies surrounding the education and assessment of ELs. The review addressed the question concerning the relationship between ELs' second language proficiency and their performance on state standardized tests. The overall aims of this chapter were to: (a) briefly overview the historical and contemporary context of standardized assessment and language policy within the United States; (b) analyze the theoretical frameworks of second language acquisition; (c) describe the stages, processes, and duration of second language acquisition for school age children; (d) analyze the divide between school

practices and current second language acquisition research; (e) examine critical issues relevant to ELs and assessment, including the cultural and linguistic validity of standardized assessments. A current gap in the literature lies with respect to bringing these different research domains together within a real world context, to more fully explore the relationship between second language proficiency and performance on state-mandated assessments, particularly with computerized performance-based assessments like PARCC and ACCESS for ELLs 2.0. Reviewing four decades of literature as it relates to the linguistic development, instruction, and assessment of ELs raises important issues concerning the equity, validity, and impact of our current practices and policies. This study attempts to empirically demonstrate the relationship between students' English language proficiency and their achievement on content-area standardized assessment for a diverse group of elementary school ELs in the State of Illinois. The overall aim is to delve deeper into what has become an unconcerned relationship for many educators, administrators, and policy-makers, providing evidence and insight to ensure a more just assessment and educational experience for ELs.

CHAPTER THREE. METHODOLOGY.

Introduction

This chapter provides a description of the methodology used to conduct the research study. Accordingly, this chapter reviews the purpose of the study, as well as the research questions and accompanying hypotheses. It provides an overview of assessments in Illinois and evaluates specific testing items within the PARCC assessment. The site selection, research design, data collection and analysis, validity, ethical considerations, and limitations of the study are delineated here. The chapter concludes by restating the goals of this research in striving to impact the understanding of educational decision-makers while contributing to a broader scholarly discussion.

Restatement of the Purpose

The study seeks to address the overarching issues concerning the processes of second language development and assessment of ELs within the context of public education in the United States. The purpose of this quasi-experimental correlational design quantitative study is to examine the relationship between ELs' second language proficiency and achievement on federally mandated standardized content-area assessments. A more clearly developed empirical understanding of the relationship between English language proficiency and grade level content-area achievement, as measured by performance-based standardized tests, will help to inform both legislators and educators as to assessment practices and policies that are most equitable and valid for ELs and their school communities.

Deemed low-stakes, non-immediate, and lacking individual impact for children, state assessments can have immediate high-stakes consequences for states, districts, schools, and

communities (Menken, 2008; Trumbull, Basterra, & Solano-Flores, 2011). This system of high-stakes accountability through standardized assessment determines which schools and students are labeled successful, and which are defined as failing. It specifies the allocation of financial resources and supports, even deciding which communities are capable of self-governing their own schools. These broader consequences shape the practices and policies experienced by children every day. The fallout of high-stakes assessments impacts the educational models, programming, curriculum, classroom make-up, and instructional practices of all students; however, it disproportionately affects culturally and linguistically diverse learners (Menken, 2008; Trumbull, Basterra, & Solano-Flores, 2011). This in addition to the more personalized impact of this deficit-based approach to education that serves to further separate students from their linguistic, cultural, and community-based identities (Valenzuela, 2002, 2005; Menken, 2008). Unintended consequences of this misalignment and system of high-stakes testing for ELs include: increased dropout rates, decreased graduation rates, higher rates of younger students taking the General Education Diploma (GED) instead of completing high school, increased numbers of low performing students being held back in grades before testing, and higher numbers of suspensions and expulsions of low performing students immediately prior to testing (Amrein & Berliner, 2002; Menken, 2008). This combination of invalid assessment practices and a culture of high-stakes accountability create a unique educational context for ELs, as well as, the districts, schools, classrooms, and communities which serve them.

Restatement of the Research Questions and Hypotheses

The overall research question guiding this study is:

How does an ELs' second language proficiency influence his/her achievement on federally mandated standardized content-area assessments?

The subsequent research questions and hypotheses for this study were as follows:

Research question 1. What is the relationship between English language proficiency and achievement on grade level content-area standardized assessments in ELA and mathematics?

Hypothesis 1. The more advanced the student's level of language proficiency in English, the higher their reading achievement.

Hypothesis 2. The more advanced the student's level of language proficiency in English, the higher their mathematics achievement.

Research question 2. To what extent does English language proficiency in reading, writing, speaking, and listening influence achievement on grade level content-area standardized assessments in ELA and mathematics?

Hypothesis 3. A student's English language proficiency in reading and writing will impact their reading achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.

Hypothesis 4. A student's English language proficiency in reading and writing will impact their mathematics achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.

Research question 3. What is the threshold English language proficiency for ELs to meet/exceed standards on federally mandated standardized assessments in ELA and mathematics?

Hypothesis 5. According to Illinois school code ELs are defined as proficient at a composite ACCESS score of 4.8 or beyond, and should thus be able to demonstrate proficiency on grade level content-area assessments.

Research question 4. What is the impact of other student demographic factors - race/ethnicity, socioeconomic status - in relation to English language proficiency for the achievement of ELs on federally mandated standardized assessments in ELA and mathematics?

Hypothesis 6. Controlling for other variables, ELs identifying with a race/ethnicity of White will achieve higher levels of reading and mathematics proficiency.

Hypothesis 7. Controlling for other variables, ELs with full pay lunch status will achieve higher levels of reading and mathematics proficiency.

Participants and Site Selection

In 2019, the State of Illinois had 1,984,519 students enrolled in its public schools (Illinois Report Card, 2019). 12.1% were identified as ELs, totaling over 240,000 students. In examining the racial and ethnic diversity of the state, nearly half, 47.6%, of students identify as White. Approximately a quarter of students, 26.4%, as Hispanic, and 16.7% as Black. Asian students make-up 5.1% of the population, while those identifying as Two or More Races are 3.8%, 0.3% as American Indian , and 0.1% as Pacific Islander. Approximately half of students within the

state, 48.8%, come from low income households, 2% are homeless. 16% of students are supported in special education with an Individualized Education Program (IEP). Average daily attendance for the state is 94%, a figure that has held consistent for the past five years. The state has a 7% mobility rate, a decline from 12% three years earlier. On average, the state spends $8,172 per-pupil in instructional expenditures and $13,764 in operational expenditures (Illinois Report Card, 2019).

To provide a more localized context, the community in which the school district and student population studied reside is a suburb bordering a large urban environment. According to the 2010 U.S. Census there was a population of 64,784 residents. In 2010, the demographic make-up of the community was 60.3% White, 25.5% Asian, 8.8% Hispanic, 7.3% Black, 0.2% Native American, and 0.02% Pacific Island, and 6.7% identifying as Other. This information does not reflect the significant demographic shifts that have transpired within the community over the past decade. An influx of immigrant and refugee populations, as well as, the departure of many White residents, particularly those of Jewish faith, have transformed the community and will be more accurately reflected in the 2020 Census. The per capita income of residents is $32,169, with a median home value of approximately $300,000. With respect to its educational system, the community is unique in that it has five distinct elementary school districts that feed into one high school district.

The participants for this study include 119 third, fourth, and fifth grade ELs in a small suburban Illinois school district during the 2017-2018 academic school year. The demographics for this first-ring suburban school district are as reported by the Illinois State Board of Education via the 2018-2019 Illinois State Report Card. The district serves just under two thousand

students in four schools kindergarten through eighth grade. The district operates with $25,686,912 in annual financial resources, funded at 103% of the state adequacy level. This equates to $9,842 per-pupil in instructional expenditures and $15,796 in operational expenditures. According to the State's evidence based funding model, the district is assigned Tier 4, receiving minimal additional financial assistance. With respect to racial and ethnic demographic information, 36.8% of students identified as White, 32.4% Asian, 17.3% Hispanic, 8.3% Black, 4.9% Two or More Races, 0.2% American Indian, and 0.1% as Pacific Islander. Additionally, the district serves a large Middle Eastern population, with many self-identifying as White on demographic forms (Public Radio International, 2017). Just over half of the students in the district, 50.1%, come from low-income households, with 0.8% identifying as homeless. 16% of students are supported in special education with an IEP. The district's 95.4% student attendance rate is slightly above the state average of 94%; while student mobility is one percent higher than the state average at 8%. 23.0% of students in the district are identified as ELs, compared to 12.1% of students within the State of Illinois. The percentage of ELs is higher at the elementary level (K-5) at just under thirty percent. Over the past five years the district has experienced a nearly forty percent increase in the number of ELs attending school within its boundaries.

This specific group of students was selected for inclusion in the study as they are identified as ELs in the elementary grade levels mandated to take the state's content area assessments. The 119 third, fourth, and fifth graders are all students within the school district in which I was a Principal at one of the elementary buildings. This point of access provided the opportunity to compile detailed datasets while ensuring the privacy and confidentiality of

participants. The demographic make-up of this student population is: 42.9% Asian, 36.1% White, 15.1% Hipsanic, 4.2% Black, 0.008% Native American and Multiracial. As a measure of socioeconomic status, nearly two thirds of these students, 65.5%, qualified for free lunch. An additional 10.1% qualified for reduced lunch, and approximately one-quarter, 24.4%, were full pay. There are 27 different languages spoken within this student group. 21.8% of ELs in this population identify Urdu as their native language, Spanish (15.1%) was second, with Assyrian (13.4%), Arabic (11.8%), and Tagalog (5.9%) rounding out the top five.

Research Design

The relationship between the ACCESS and PARCC scores for a small, yet diverse, group of third through fifth grade ELs was investigated in this quasi-experimental correlational design quantitative study. A correlational design is utilized for this study as a means to relate two variables. The independent variable, English language proficiency, is defined as a student's composite scale score and accompanying proficiency level on the ACCESS for ELLs 2.0 assessment. The dependent variable, achievement on grade level content-area standardized assessment, is defined as a student's overall scale score and accompanying proficiency level on the PARCC assessment for both ELA and mathematics. This study employs various statistical tests - linear correlation, Chi-Square test of independence, and linear regression models - to measure the degree of association and strength of relationship amongst the variables to test the hypotheses and answer the research questions.

Data Collection

Data was obtained from the 2017-2018 ACCESS for ELLs 2.0 assessment, 2018 PARCC Assessment for ELA and mathematics, and student registration information for the 2017-2018

school year. Student information was aggregated into a spreadsheet by the district's Data Services Specialist. This central office position is responsible for compiling and reporting all of the district's data within the state's Student Information System. The data was compiled using random identification numbers to ensure that student identity and privacy are concealed and fully protected. The data - student identifier, grade level, race/ethnicity, home language, lunch status, composite ACCESS proficiency level and scale score, ACCESS reading proficiency level and scale score, ACCESS writing proficiency level and scale score, ACCESS listening proficiency level and scale score, ACCESS speaking proficiency level and scale score, PARCC reading proficiency level and scale score, PARCC mathematics proficiency level and scale score - were recorded in tabular form in an Excel spreadsheet and imported into IBM's SPSS Statistics software. In accordance with the policies of DePaul University's Institutional Review Board (IRB) this study was deemed to be non-reviewable. From an IRB perspective, the study does not involve human subjects since de-identified data sets cannot be traced back to the original subjects.

Overview of Assessments in Illinois

ESSA requires states to adopt English language development standards aligned to academic standards, as well as, administer assessments for both English language and academic content area proficiency (Menken, 2008; Lyons & Dadey, 2017). At the time of this study, 2017-2018 school year, the State of Illinois utilized ACCESS for ELLs 2.0 as the assessment of English language proficiency and the PARCC assessment for content-area proficiency in English language arts/literacy (ELA) and mathematics.

The Partnership for Assessment of Readiness for College and Careers (PARCC) Assessment

PARCC was a computer-based standardized assessment administered to third through eighth grade and high schoolers in the State of Illinois, and other PARCC consortium states, from 2015 through 2018. The overall assessment was made up of multiple tests, with the precise number varying by grade level and dependent upon the year of administration, for the domains of ELA and mathematics. The sample of students for this study took the PARCC exam in 2018. That assessment was made up of three ELA and four mathematics tests (PARCC Assessment, 2019).

In 2010, the PARCC consortium, consisting of 24 states and the District of Columbia, was awarded funding as part of the federal government's *Race to the Top* program. *Race to the Top* sought to create and implement a standard set of K-12 ELA and mathematics assessments based on the Common Core State Standards. According to Pearson, developer of PARCC, more than two thousand educators, researchers, psychometricians and other stakeholders collaborated to develop the annual assessment. The two overarching goals of the assessment were to evaluate the preparedness of students for college and technical courses, and to define the knowledge and skills students should demonstrate at any particular grade level (PARCC Assessment, 2019).

For the 2018-2019 school year the State of Illinois left the consortium and replaced PARCC with the Illinois Assessment of Readiness (IAR). After an initial vendor dispute, Pearson maintained the contract to continue to provide the annual content-area proficiency assessment. Accordingly, IAR utilizes the same test content, construction, and format as the PARCC assessment. IAR has about one third less overall test items than PARCC (Illinois State

Board of Education, 2019). From a student perspective the test itself presents the same. Performance levels, descriptors, and cutoff scores remained the same between the two assessments; this continuity was intentional and allowed the state/districts to track student performance year-to-year.

Evaluating Student Performance

Student's receive a scale score for both ELA and mathematics, ranging from 650 to 850 (PARCC Assessment, 2019). This scale score is then associated with one of five proficiency levels: 650-699 level 1 (did not yet meet expectations), 700-724 level 2 (partially met expectations), 725-749 level 3 (approached expectations), 750-809 level 4 (met expectations), and 810-850 level 5 (exceeded expectations). *Figure 8* shows the scoring scale and corresponding performance levels as presented on an individual student's score report.

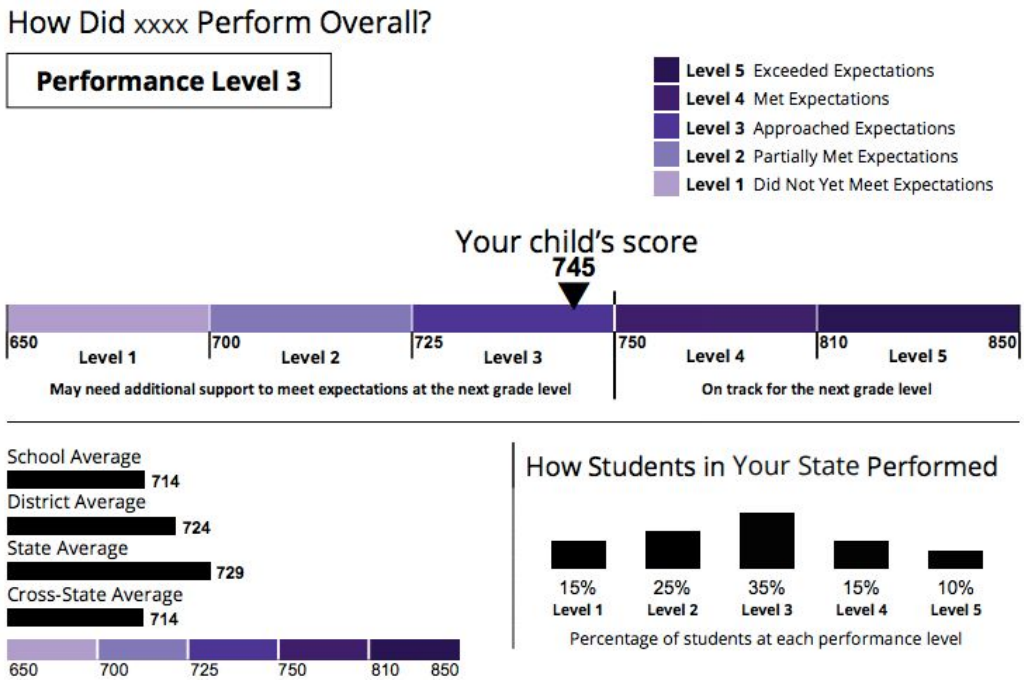


Figure 8. PARCC Sample Student Scoring Report.

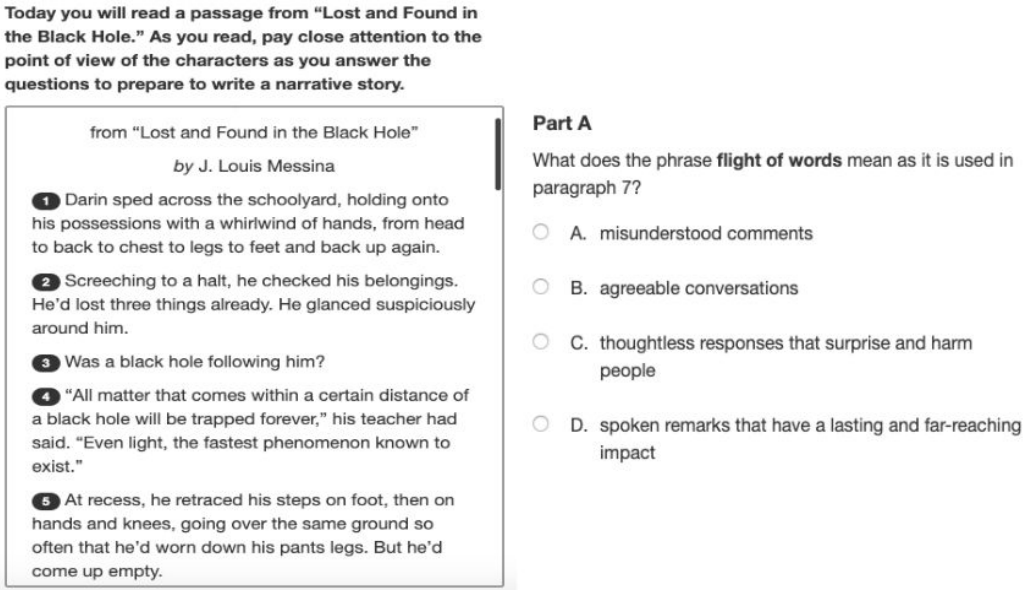
PARCC scores are intended to summarize student achievement towards the Illinois Learning Standards for English language arts and mathematics, providing information to students, families, educators, schools, and districts that indicate areas of mastery and ongoing improvement. The aim of such data is to inform practices, resources, and supports at the classroom, school, and district level, as well as, provide additional information about which students, both individual and groups, are achieving academically. The State of Illinois, currently uses these assessment scores, weighted at seventy percent, to determine a school's rating. Schools within the state are rated as one of four summative designations - lowest performing, under performing, commendable, and exemplary (Illinois Report Card, 2019).

English Language Arts/Literacy

The ELA portion of the PARCC exam assesses both reading and writing skills, with each domain weighted as fifty percent of the overall ELA score. The subdomains of the reading portion are literary text, informational text, and vocabulary. Writing expression and knowledge and use of language conventions are the subdomains for writing. Though weighted equally and assessed together, reading comprehension makes up a majority of the test items, underlying the written portions as well. Students are asked to read and respond to passages from real texts, both fiction and nonfiction, as well as multimedia video and audio. ELA items are made up of both on-demand selected and constructed responses - multiple choice, drag and drop, sequence events, and extended written response (PARCC Assessment, 2019). Critical issues for ELs with respect to PARCC ELA are consistent with those outlined in the research - construct irrelevance due to linguistic demands, appropriateness and relevance of text selections, effectiveness of accommodations, evaluator bias of scoring written work, and overall linguistic complexity of the

assessment and test directions (Abedi, 2002, 2006, 2011; Abedi, Hofstetter, & Lord, 2004; Durán, 2011; Menken, 2008).

The extent to which these challenges impact the validity and reliability of assessment results for ELs is most clearly illustrated through actual assessment artifacts. *Figure 9* is an example of a literary analysis task for fifth grade. It is a multiple-choice format on-demand selected response item.



Today you will read a passage from "Lost and Found in the Black Hole." As you read, pay close attention to the point of view of the characters as you answer the questions to prepare to write a narrative story.

from "Lost and Found in the Black Hole"

by J. Louis Messina

1 Darin sped across the schoolyard, holding onto his possessions with a whirlwind of hands, from head to back to chest to legs to feet and back up again.

2 Screeching to a halt, he checked his belongings. He'd lost three things already. He glanced suspiciously around him.

3 Was a black hole following him?

4 "All matter that comes within a certain distance of a black hole will be trapped forever," his teacher had said. "Even light, the fastest phenomenon known to exist."

5 At recess, he retraced his steps on foot, then on hands and knees, going over the same ground so often that he'd worn down his pants legs. But he'd come up empty.

Part A

What does the phrase **flight of words** mean as it is used in paragraph 7?

A. misunderstood comments

B. agreeable conversations

C. thoughtless responses that surprise and harm people

D. spoken remarks that have a lasting and far-reaching impact

Figure 9. Grade 5 PARCC/IAR ELA Assessment Item.

This singular passage is 805 words in length, and places significant linguistic demands upon ELs in a way that it does not for native English speakers. Though this reading and its accompanying questions are intended to assess comprehension of a literary text, the scientific nature of the article introduces several impacting variables. The elements of passive voice, idiomatic language, and scientific vocabulary present construct irrelevant variables for ELs in ways they do not for other students. Many of the difficult terms are those having different meanings in everyday language versus technical language (e.g. matter) or those that do not refer to scientific

concepts but ensure precision in scientific reasoning (e.g. likely, significant) (Trumbull & Solano-Flores, 2011; Wellington & Osborne, 2001). Furthermore the sentence and text complexity, absent a meaningful context for language usage in standardized assessment, can interfere with the comprehension and performance of ELs (Ferguson, 1985; Trumbull & Solano-Flores, 2011).

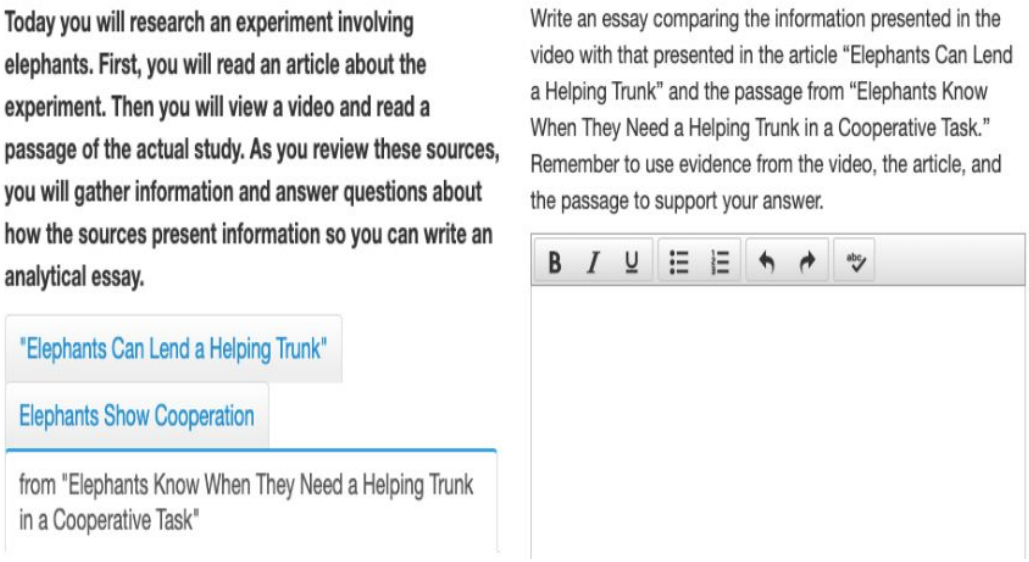


Figure 10. Grade 8 PARCC/IAR ELA Assessment Item.

Figure 10 is an extended length on-demand constructed response item from an eighth grade research simulation task. This assessment task requires that students decipher and interpret a scientifically based article and video. They must then compare and contrast the two sources of information, using key ideas and details supported by evidence, to establish their point of view. This task of language production is complex and challenging for all students, but particularly so for students learning the English language. As can be seen by this assessment item, a student's writing ability is a significant factor upon their overall literacy score. Writing scores impact fifty percent of a student's PARCC ELA assessment. Furthermore, responses coded "Response is not

written in English" are deemed unscorable, with no points awarded. Hence an even greater discrepant impact, resulting from both format and scoring of this assessment, affects ELs in ways that it does not native English-speaking students.

Figure 11 illustrates the differential performance between ELs and the overall student population taking the PARCC ELA assessment for the 2017-2018 school year. There is a twenty-six percentage point gap between ELs and the entire student population with regard to having achieved at the "met/exceeded" performance levels. Additionally, the percentage of ELs performing at the "did not meet" level is nearly two and a half times the percentage for all students. Only at fourth grade did a single percentage of ELs reach the "exceeded" performance level, while the average was six percent for all students throughout the state. Additionally, only four percent of fifth grade ELs "met" the proficiency standards, leaving ninety-six percent that failed to score at a level deemed proficient headed into middle school. This compared to thirty-six percent of all fifth grade students who "met/exceeded" standards. A macro perspective of the data illustrates vastly different levels of performance on this assessment for ELs as compared to the general student population.

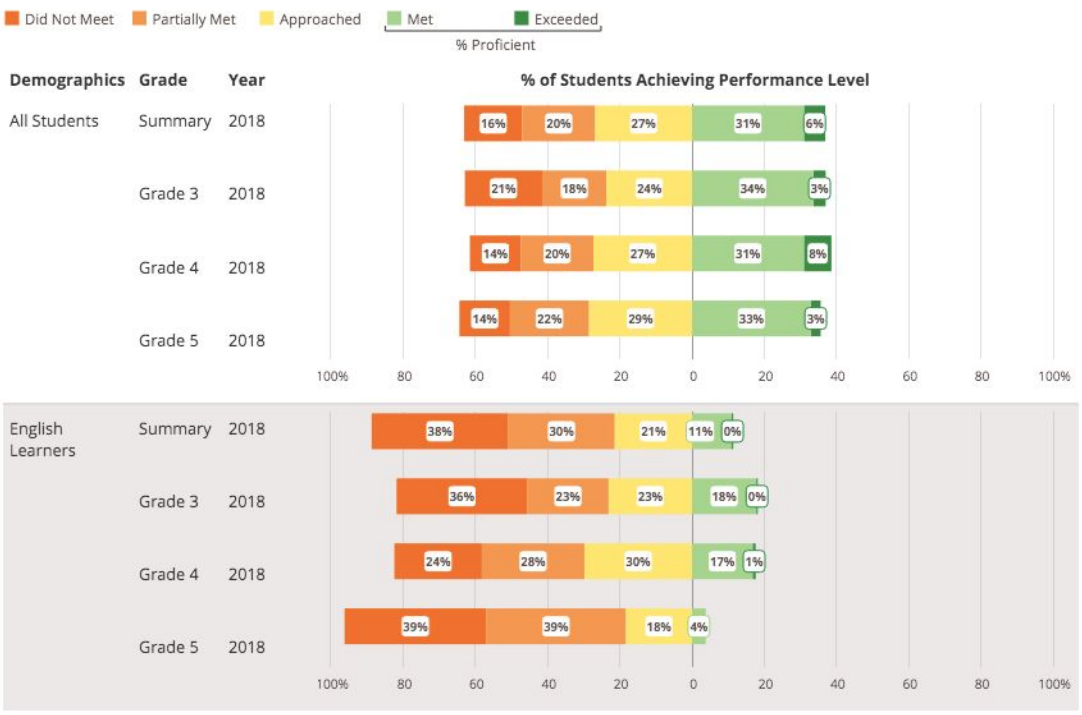


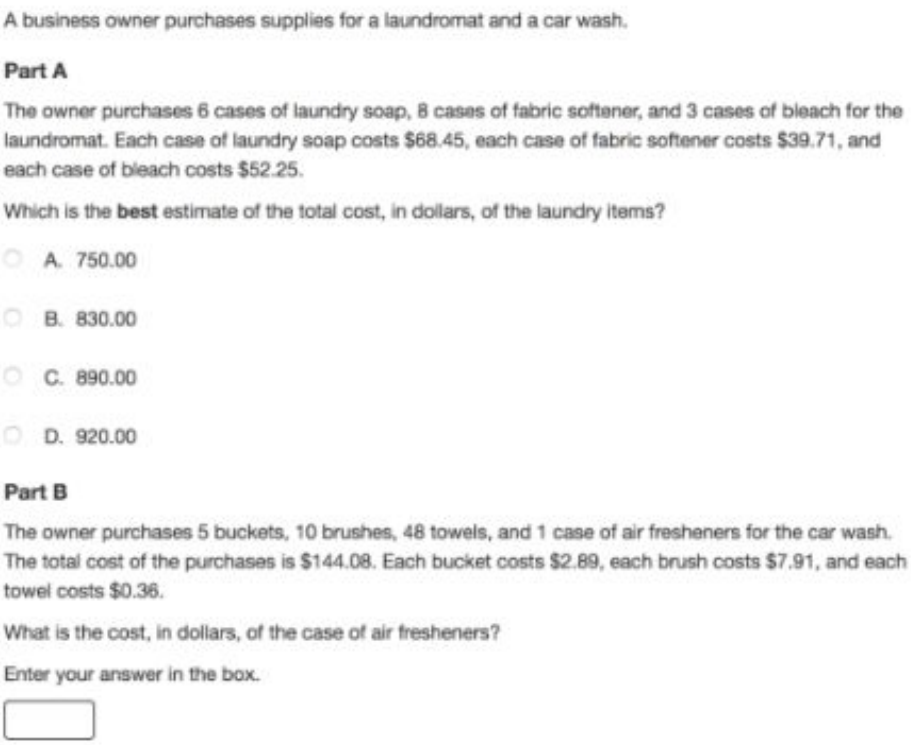
Figure 11. Percentage of "English Learners" Versus "All Students" at Each Performance Level on the 2018 PARCC ELA Assessment.

Mathematics

The mathematics portion of the PARCC exam assesses the subdomains of major content, additional and supporting content, expressing mathematical reasoning, and modeling and application. Students are asked to solve multi-step math problems that require reasoning within a real-world context. Mathematics items are made up of both on-demand selected and constructed responses - multiple choice, short and multi-step written responses (PARCC Assessment, 2019). Critical issues for ELs with respect to PARCC are consistent with those outlined in the research - construct irrelevance due to linguistic demands, appropriateness and relevance of text selections, effectiveness of accommodations, evaluator bias of scoring written

work, and overall linguistic complexity of the assessment and test directions (Abedi, 2002, 2006, 2011; Abedi, Hofstetter, & Lord, 2004; Durán, 2011; Meken, 2008).

The extent to which these challenges impact the validity and reliability of assessment results for ELs in mathematics is most clearly illustrated through actual assessment artifacts. *Figure 12* is a seventh grade example from a task assessing concepts, skills, and procedures. Part A is a multiple-choice format on-demand selected response item. Part B is a short-length on-demand constructed response item.



A business owner purchases supplies for a laundromat and a car wash.

Part A

The owner purchases 6 cases of laundry soap, 8 cases of fabric softener, and 3 cases of bleach for the laundromat. Each case of laundry soap costs $68.45, each case of fabric softener costs $39.71, and each case of bleach costs $52.25.

Which is the **best** estimate of the total cost, in dollars, of the laundry items?

A. 750.00

B. 830.00

C. 890.00

D. 920.00

Part B

The owner purchases 5 buckets, 10 brushes, 48 towels, and 1 case of air fresheners for the car wash. The total cost of the purchases is $144.08. Each bucket costs $2.89, each brush costs $7.91, and each towel costs $0.36.

What is the cost, in dollars, of the case of air fresheners?

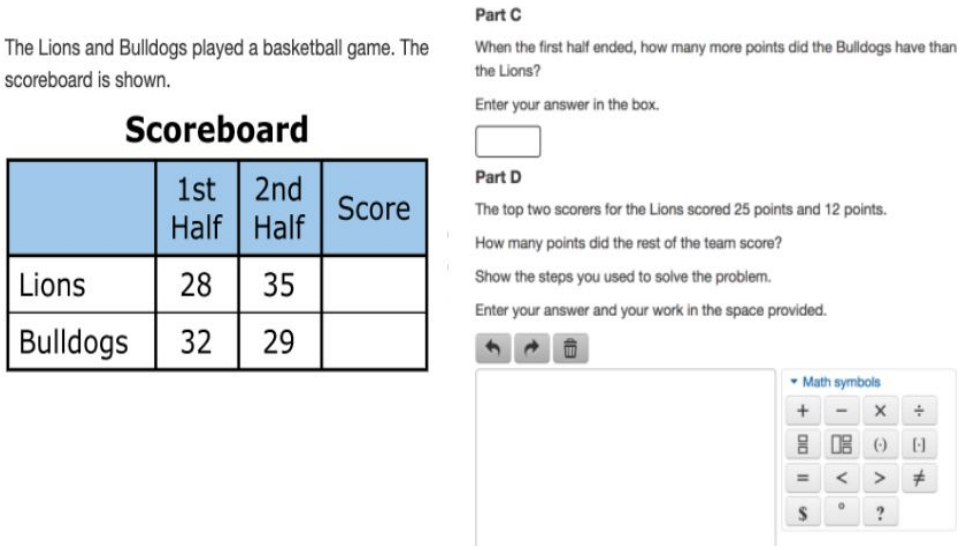
Enter your answer in the box.

Figure 12. Grade 7 PARCC/IAR Mathematics Assessment Item.

This assessment item, purportedly assessing only mathematical skills and concepts, is highly dependent on language and literacy skills. It is a heavily text-based word problem that requires students to decipher the language and register of mathematics to determine the precise

calculations needed to deduce the correct answer. *Figure 12* exemplifies both the linguistic and culture burden underpinning this mathematics problem for ELs. In Part A the use of the word "case" is unnecessarily problematic as it has multiple meanings depending on context - situational ("a case of mistaken identity"), legal ("a libel case"), medical ("15,000 cases of the Coronavirus)", as well as its intended use here as a container designed to hold something. Even with its intended use there are multiple interpretations depending upon the part of speech. As a noun case is a container, versus its use as a verb meaning to enclose ("steel cased") or to inspect ("casing the hideout"). Absent a more full and meaningful context for language in standardized assessment, the wording and sentence structure can interfere with the comprehension and performance of ELs in ways it does not for native speakers of English (Ferguson, 1985; Trumbull & Solano-Flores, 2011). This problem exemplifies the linguistic complexity of instructional assessments as the major factor contributing to the performance gap between ELs and their monolingual English speaking peers (Abedi, 2006, 2011; Abedi, Hofstetter, & Lord, 2004; Solano-Flores & Trumbull, 2003; Solano-Flores, 2011). Additionally, the "real-world" application of this assessment item is grounded in stark socioeconomic differences. Again in *Figure 12*, the operating construct of the problem is the student as owner of the laundromat. This construct is problematic for many students who, for socioeconomic or cultural reasons, may utilize laundromats to clean their clothes. First, the problem presumes a laundromat's responsibility of providing cleaning supplies to customers; whereas many students would have the lived experience of purchasing those items elsewhere at a lesser cost and bringing them to the laundromat to make use of the machines. Furthermore, the price points provided in the problem vary significantly from any real-life experience that students would have in purchasing these

items and negates any such actual experience. As such, students of lower socioeconomic status, of which ELs make up a greater proportion, may have lived experiences that create broader equity concerns in allowing all students to respond to these questions equally and as intended.


The Lions and Bulldogs played a basketball game. The scoreboard is shown.

Scoreboard

	1st Half	2nd Half	Score
Lions	28	35	
Bulldogs	32	29	

Part C

When the first half ended, how many more points did the Bulldogs have than the Lions?

Enter your answer in the box.

Part D

The top two scorers for the Lions scored 25 points and 12 points.

How many points did the rest of the team score?

Show the steps you used to solve the problem.

Enter your answer and your work in the space provided.

Math symbols

Figure 13. Grade 3 PARCC/IAR Mathematics Assessment Item.

Figure 13 is a third grade example from a task for expressing mathematical reasoning. Part C is a short-length on-demand constructed response item. Part D is an on-demand constructed response item of medium to extended length. In this assessment task a student must again be able to decipher language to determine the mathematical calculations needed to arrive at the correct answer. Moreover, this item asks students to write out the steps involved in solving the problem and to explain their thinking. "Showing the process for attaining an answer involves language to explain the steps involved. As such, an EL student must not only be able to comprehend the math register, they must be able to produce it as well" (Menken, 2008, p. 79). This item is reflective of broader shifts - more complex language usage and demonstrating a student's ability to show how they arrived at an answer - in performance-based assessments.

Similar to the previous example, this item poses additional concerns of cultural validity, as a more detailed understanding of the American sport of basketball provides an inherent advantage when responding to the questions. The combination of these factors create a test experience riddled with construct irrelevant variables that result in a significantly discrepant testing experience for ELs than their English-speaking peers.

Figure 14 illustrates the differential performance between ELs and the overall student population taking the PARCC mathematics assessment for the 2017-2018 school year. There is a twenty percentage point gap between ELs and the entire student population with regard to having achieved at the "met/exceeded" performance levels. Additionally, the percentage of ELs performing at the "did not meet" level is nearly double the percentage for all students. Only a single percentage of ELs reached the "exceeded" performance level, while the average was five percent for all students throughout the state. Additionally, only five percent of fifth grade ELs "met" the proficiency standards, leaving ninety-five percent that failed to score at a level of deemed proficient headed into middle school. This compared to thirty-one percent of all fifth grade students who "met/exceeded" standards. A macro perspective of the data illustrates vastly different levels of performance on this assessment for ELs as compared to the general student population.

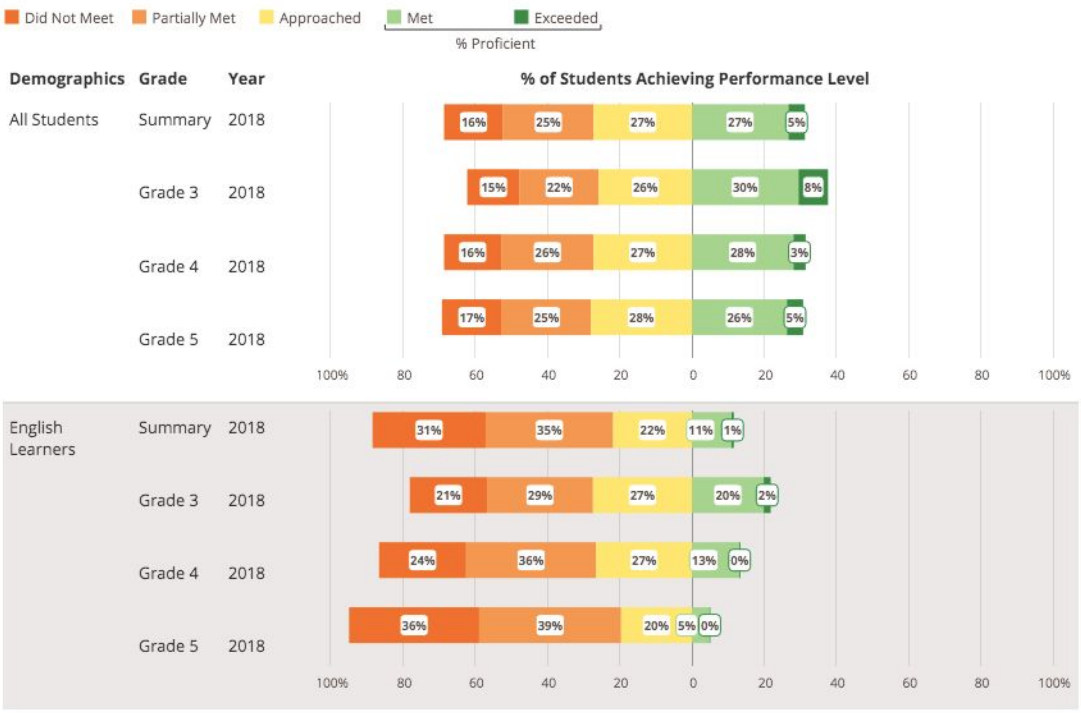


Figure 14. Percentage of "English Learners" Versus "All Students" at Each Performance Level on the 2018 PARCC Math Assessment.

ACCESS for ELLs 2.0

The ACCESS for ELLs 2.0 assessment was developed by the WIDA Consortium, formerly known as World-Class Instructional Design and Assessment. WIDA is an educational consortium currently made up of 40 states, including Illinois, as well as territories, federal agencies, and hundreds of international schools. WIDA designs and implements English language development proficiency standards and assessments for kindergarten through twelfth grade ELs. Additionally, WIDA provides professional learning opportunities for educators and conducts research on instructional and assessment practices for ELs (WIDA, 2019).

ACCESS is a computer-based, adaptive test that assesses students' language in four domains: listening, reading, speaking, and writing. The assessment is given annually to ELs in

first through twelfth grades to assess English language proficiency, and can be administered in a group or individual setting (WIDA, 2019). Kindergarten students take a paper-based test administered individually and in a game-like interactive format.

The ACCESS test is made up of four parts - reading, listening, speaking, and writing - and requires 170 minutes of total test time. The test is quasi-adaptive and responds to the performance of individual students, routing students towards easier or more challenging content based on their performance. The listening and reading domains are the adaptive sections and must be assessed first. Students are then grouped into one of three levels of tiers for the speaking and writing portion of the assessment based upon their initial performance. There are no tier changes in the assessment after this initial placement. According to WIDA, the test is designed for increased student engagement and has built-in accessibility and accommodations features to fit more individualized learning needs (WIDA, 2019). Unlike the PARCC assessment, proctors of ACCESS are required to complete an online training course prior to being certified to administer the test.

ACCESS items are made up of both on-demand selected and constructed response items. Selected response items are in multiple choice format with students asked to select the "correct" answer. Constructed response items require an independent written or spoken response. For the writing test, students type their answers into the box, while on the speaking test they speak a response into a microphone (WIDA, 2019). Per ESSA, schools within Illinois are mandated to give the ACCESS to ELs annually within an established timeframe of early January to mid-February.

Evaluating Student English Language Proficiency

Student's receive a scale score and proficiency level for each of the four language domains - listening, speaking, reading, and writing - as well as, additional domains that factor these initial domains together - oral language, literacy, and comprehension. Students are assigned an overall composite scale score and proficiency level. Scale scores are between the ranges of 100 and 600 for each of the domains. This scale score is then associated with one of six proficiency levels based on the six WIDA English language development levels discussed earlier in this review - level 1 entering, level 2 emerging, level 3 developing, level 4 expanding, level 5 bridging, and level 6 reaching. The oral language domain is a combination of the listening and speaking tests, weighted together at fifty percent each. The literacy domain is a combination of the reading and writing tests, weighted together at fifty percent each. The comprehension domain is a combination weighting the reading test at seventy percent and the listening test at thirty percent. A student's overall composite score is weighted at thirty-five percent reading, thirty-five percent writing, fifteen percent listening, and fifteen percent speaking. In the state of Illinois, an overall composite score of 4.8 is the threshold for ELs to be reclassified and to no longer qualify for additional language support (WIDA, 2019). *Figure 15* shows the scale score and corresponding proficiency levels as presented on an individual student's score report.

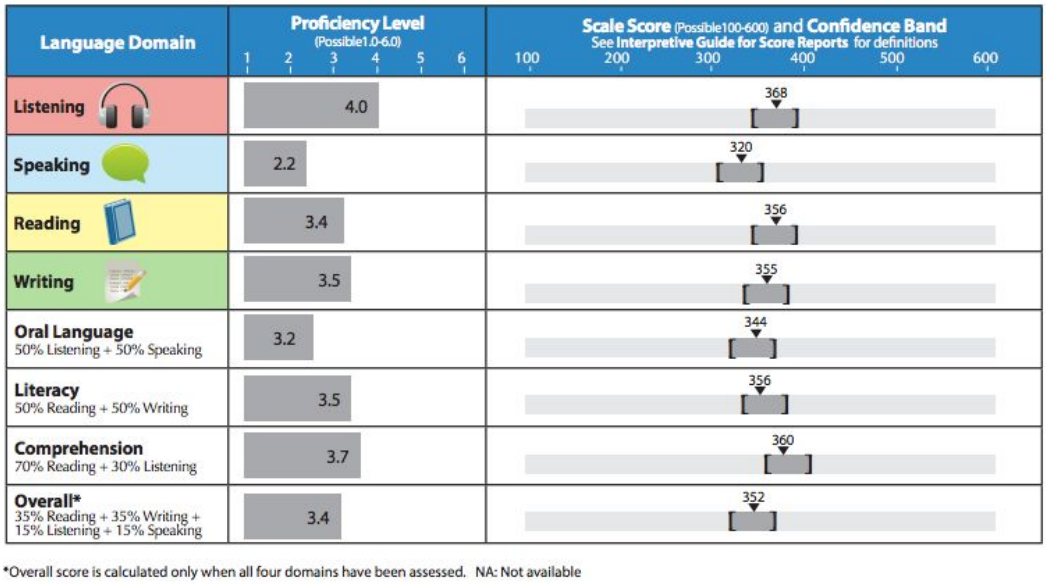


Figure 15. Sample Student Scoring Report for ACCESS for ELLs 2.0.

ACCESS scores are intended to provide greater detail on an individual student's level of English language proficiency relative to the four domains assessed. Domain scores are scaled vertically to allow for comparisons within a domain over time (WIDA, 2019). Accordingly, ACCESS is an assessment that allows for tracking student growth and language development over time. The proficiency levels alignment to the six WIDA performance levels also coordinate within the broader WIDA language development framework and "can-do descriptors" that were described in the literature review. The WIDA framework provides students, families, and educators with clear and actionable descriptors of a student's readiness for instruction. School, district, and state reports allow for broader analyses to inform practices, resources, and supports at the classroom, school, and district level, as well as, provide additional information about which students, both individual and groups, are achieving linguistic growth relative to the four assessed domains (WIDA, 2019).

Despite the fact that ACCESS is an assessment designed for ELs to determine their English language proficiency, the critical issues are consistent with those previously discussed in this review of the research - construct irrelevance due to linguistic demands, effectiveness of accommodations, changes made to the scoring, differential weighting of language domains, and overall difficulty of the assessment (Abedi, 2002, 2006, 2011; Abedi, Hofstetter, & Lord, 2004; Durán, 2011; Meken, 2008). Construct irrelevance for the ACCESS test derives primarily from the difficulty of separating language proficiency from academic language, and overall validity of attempting to convey application of language through a standardized assessment. Additional issues arise with the application of data at the state, district, school, and classroom level. ACCESS data alone should not be the sole factor in the redesignation of ELs. This information should be utilized alongside multiple other measures - teacher evaluation, observation, parent input, academic achievement on content-area standardized assessments - to determine placement, programming, and additional language and academic supports for students (Trumbull, Basterra, & Solano-Flores, 2011). ELs are an incredibly diverse and complex student population with varied needs; ensuring a meaningful and responsible alignment between assessment, data, and decision-making must be a top priority for all educators and policy-makers.

Instrumentation

This study uses three different instruments for data analysis. Each was discussed in detail and with the inclusion of critical issues for ELs in the previous section of this chapter, each is briefly described as follows. The ACCESS for ELLs 2.0 assessment is the instrument used to measure an EL's level of language proficiency in English. This computer-based standardized test assesses four language domains - reading, writing, speaking, and listening - providing a scale

score (100-600) and accompanying proficiency level (1.0-6.0). The PARCC assessment is the instrument utilized to measure content-area proficiency in ELA and mathematics. This computer-based standardized test assesses the subdomains of literary text, informational text, and vocabulary for reading, and writing expression and knowledge and use of language conventions for writing. The mathematical subdomains assessed are major content, additional and supporting content, expressing mathematical reasoning, and modeling and application. PARCC performance is measured by a scale score for both ELA and mathematics, ranging from 650 to 850. This scale score is then associated with one of five proficiency levels: 650-699 level 1 (did not yet meet expectations), 700-724 level 2 (partially met expectations), 725-749 level 3 (approached expectations), 750-809 level 4 (met expectations), and 810-850 level 5 (exceeded expectations). The final instrument used for this study is student registration information from the 2017-2018 school year. This includes the demographic information that families report as part of their initial paperwork upon registering within the district at the beginning of the school year.

Data Analysis

As was previously discussed, this study employs various statistical tests - linear correlation, Chi-Square test of independence, and linear regression models - to measure the degree of association and strength of relationship amongst the variables to test the seven hypotheses and answer the four research questions.

Linear correlation was the statistical test utilized to answer the first research question and test its underlying hypotheses. Researchers use correlation as "a way of understanding the association between two variables" (Abbott, 2011, p. 337). Pearson's Correlation Coefficient, *r*,

is used to measure the relationship between the two interval level variables of the data. With respect to the first part of the initial research question, composite ACCESS scale scores are correlated with PARCC reading scale scores to determine the strength and direction of the correlation. Similarly, for the second part of the research question, composite ACCESS scale scores are correlated with PARCC mathematics scale scores to determine the strength and direction of the correlation. Scattergrams serve to visually represent the data. The Hypothesis Test for Pearson's *r* is used as the first procedure for determining the significance of the relationship between the two study variables. If the *r* value exceeds the tabled critical value, then we can conclude the relationship is statistically significant (Abbott, 2011). Comparison to the *T* table of values is a second method for identifying critical values and determining the significance of the relationship between the two variables. This is done by transforming the *r* value into a *t* ratio, that can then be compared against the critical values of *T* in the *Exclusion Values for the T Distribution* table (Appendix B). If the *t* score exceeds the critical value then the relationship is significant (Abbott, 2011). Effect size is the third and final technique used to judge the significance and strength of the correlation. The coefficient of determination, r^2, can be used to assess the strength of the relationship, and refers to the amount of variance in one variable as explained by the other (Abbott, 2011). Abbott (2011) defines this as "the percentage of the variance in one variable (outcome) contributed by another (predictor)" (p. 356). The r^2 value is the explained variance, thus leaving the remaining percentage as the unexplained variance caused by other variables. Cohen (1988) established the following conventions for determining the effect size of r^2: small effect size, $r^2 = 0.01$ ($r = 0.10$); medium effect size, $r^2 = 0.09$ ($r = 0.30$); large effect size, $r^2 = 0.25$ ($r = 0.50$). These three approaches are used to determine the

strength of the relationship between English language proficiency and achievement on grade-level content-area standardized assessments in response to the first research question. Cross-tabulation analyses are performed to further analyze the connection between these two study variables. Furthermore, the Chi-Square test of independence was used to detect patterns that may indicate relatedness between the two variables. As stated in Abbott (2011) "the Chi Square procedure statistically analyzes the differences among the data in contingency tables to determine whether the patterns of difference are different enough to be considered statistically significant" (p. 454). In applying this test, if we reject the null hypothesis, stating that there is complete independence and no relationship among the variables, then we can demonstrate a dependent relationship between English language proficiency and achievement on content-area assessments.

Linear correlation will again be utilized to address the second research question and its underlying hypotheses. In addition, multiple linear regression will be employed to answer the research question and test its accompanying hypotheses. Multiple linear regression is "a process that creates a model for predicting values of an outcome variable from predictor variables, along with a way of explaining the variance in an outcome variable" (Abbott, 2011, p. 430). This regression analysis will again allow for Pearson's Correlation Coefficient, r, to be used to measure the relationship between the interval level variables of the data, correlating the strength and direction of the relationship between the independent variables (scale score of each language domain) and the dependent variables (overall PARCC ELA and mathematics scale scores). "M[ultiple] L[inear] R[egression] can be used to pinpoint the contribution of r^2 of individual predictors. This in one of the chief contributions of MLR to research" (Abbott, 2011, pp.

431-32). In answering the second research question and testing its underpinning hypotheses, r^2 values will describe the amount of variance in one variable explained by another and can be used to help determine how much variance is explained and the significance of that variance relative to the effect size (Cohen, 1988). The standardized Beta coefficients and *t*-test are used to determine significance of the independent variables (scale score of each language domain) with PARCC ELA and mathematics scale scores as the dependent variable.

With respect to the third research question, comparing mean and median ACCESS composite scores within the ranges of PARCC academic proficiency levels are used to provide an answer to the question and test its underlying hypothesis. This treatment of the data allows for determining the overall composite ACCESS score that serves as the threshold for achieving proficiency on the PARCC ELA and mathematics assessment for our particular student population. Histograms and cross tabulation tables will visually illustrate the data.

The use of seven linear regression models will help to answer the fourth research question and test its accompanying hypotheses. This allows for the comparison of model summary changes with the introduction of additional predictor variables (Abbott, 2011). According to Abbott (2011) "these subvariables can then be understood in their relationship to the outcome variable" (p. 432). Overall, PARCC scale scores for ELA and mathematics are the dependent (outcome) variable and English language proficiency level, language domains - speaking, listening, reading, and writing, socioeconomic status, and race/ethnicity the independent variables.

Model one includes English language proficiency variables only. The PARCC scale scores for ELA and mathematics are the dependent variables, with English language proficiency

levels as the independent variable. A student's level of English language proficiency is measured using the ACCESS 2.0 composite score and defined by the accompanying WIDA proficiency level and performance descriptors. The composite ACCESS score is derived by weighting and combining the four language domains - 35% reading + 35% writing + 15% listening + 15% speaking. For the purposes of this study, ELs are divided into three levels measuring different degrees of English language proficiency - low, medium, and high - this variable is operationalized by dummy variables for each distinct level of language proficiency, as shown in *Table 4*. Per Abbott (2011) "transforming predictors to sets of subvariables can be done using dummy" variables (p. 432). A dummy variable is defined for each range of English language proficiency; for example, 1 if the student has low English proficiency, and 0 if otherwise. Students whose overall composite ACCESS score was between a 1.0 and 2.9 were identified as having low English proficiency. These proficiency levels correspond with the entering (level 1) and beginning (level 2) performance levels as identified by WIDA. Students whose overall composite ACCESS score was between 3.0 and 4.9 were identified as having medium English language proficiency. These proficiency levels correspond with the developing (level 3) and expanding (level 4) performance levels as identified by WIDA. Students whose overall composite ACCESS score was above a 5.0 were identified as having high English language proficiency. These proficiency levels correspond with the bridging (level 5) and reaching (level 6) performance levels as identified by WIDA. Students with low English language proficiency served as the reference category for the first model.

Model two includes only the use of subscores for the four assessed language domains - speaking, listening, reading, and writing. The PARCC scale scores for ELA and mathematics are

the dependent variables, with the scale scores for each of the four assessed language domains - speaking, listening, reading, and writing - as the independent variables. A student's level of English language proficiency in each domain is measured using the ACCESS 2.0 scale score for that particular domain.

Model three consists of socioeconomic status variables only. The PARCC scale scores for ELA and mathematics are the dependent variables, with socioeconomic levels as the independent variable. A student's socioeconomic status is measured using their families qualification for the National School Lunch Program. Under current eligibility guidelines to qualify for free lunch, a family of four must have an annual household income of less than $33,345. To qualify for reduced lunch, a family of four must have an annual household income of less than $47,638 (U.S. Department of Agriculture, 2019). For the purposes of this study, ELs are divided into three socioeconomic levels - free, reduced, and full pay. This variable is operationalized by dummy variables for each level of socioeconomic status as shown in *Table 4*. A dummy variable is defined for each range of socioeconomic status; for example, 1 if the student qualifies for free lunch, and 0 if otherwise. Students qualifying for free lunch served as the reference category for this model.

Model four includes race and ethnicity variables only. The PARCC scale scores for ELA and mathematics are the dependent variables, with race/ethnicity as the independent variable. A family's self-identification of demographic information on registration forms is used to define a student's race/ethnicity. As outlined in *Table 4*, this variable is operationalized through five dummy variables, one for students identifying as Asian, one for Black, one for Hispanic, one for

White, and one for Other constituting American Indian, Pacific Islander, and Two or More Races. Students identifying as Black served as the reference category.

Model five combines the first two models, with PARCC scale scores for ELA and mathematics as the dependent variables, and English language proficiency levels and the scale scores for each of the four ACCESS assessed language domains as the independent variables. There exists a degree of overlap within the model, as a weighted combination of the four language domains is used to determine the composite scale scores used to define the language proficiency classifications.

Model six merges models one, two, and three together, with PARCC scale scores for ELA and mathematics as the dependent variables, and English language proficiency levels, the scale scores for each of the four ACCESS assessed language domains, and socioeconomic levels as the independent variables.

The seventh and final model combines models one, two, three, and four together, to offer the most comprehensive view of the relationship among the study variables. The PARCC scale scores for ELA and mathematics are the dependent variables, and English language proficiency levels, the scale scores for each of the four ACCESS assessed language domains, socioeconomic levels, and race/ethnicity as the independent variables.

Table 4 outlines the definitions of the variables to be included in this analysis, demonstrating how the independent and control variables were operationalized in helping to translate the research questions and hypotheses into a research methodology and statistical modeling.

Variable	Definition of the variable
Low English Language Proficiency***	1, if the student's overall ACCESS 2.0 proficiency level score was between 1.0 and 2.9 (entering & beginning WIDA levels); 0, other
Medium English Language Proficiency	1, if the student's overall ACCESS 2.0 proficiency level score was between 3.0 and 4.9 (developing & expanding WIDA levels); 0, other
High English Language Proficiency	1, if the student's overall ACCESS 2.0 proficiency level score was above 5.0 (developing & expanding WIDA levels); 0, other
ACCESS 2.0 Listening Domain Scale Score	student's scale score for the listening domain of the ACCESS 2.0 assessment (100-600)
ACCESS 2.0 Speaking Domain Scale Score	student's scale score for the speaking domain of the ACCESS 2.0 assessment (100-600)
ACCESS 2.0 Reading Domain Scale Score	student's scale score for the reading domain of the ACCESS 2.0 assessment (100-600)
ACCESS 2.0 Writing Domain Scale Score	student's scale score for the writing domain of the ACCESS 2.0 assessment (100-600)
Free Lunch Status***	1, if the student qualifies for free lunch; 0, other
Reduced Lunch Status	1, if the student qualifies for reduced lunch; 0, other
Full Pay Lunch Status	1, if the student qualifies for full pay lunch; 0, other
Asian	1, if the family of the student self-identified as Asian on registration demographic information; 0, other
Black***	1, if the family of the student self-identified as Black on registration demographic information; 0, other
Hispanic	1, if the family of the student self-identified as Hispanic on registration demographic information; 0, other
Other	1, if the family of the student self-identified as Two or More Races, American Indian, or Pacific Islander on registration demographic information; 0, other
White	1, if the family of the student self-identified as Asian on registration demographic information; 0, other
***reference group	

Table 4. Definition of the Variables in the Analysis.

Ethical Considerations

The identity of students has been fully protected as no names shall be obtained as part of the data collection process. The data provided to the researcher was compiled using random identification numbers, with the researcher having no access to identifiable information. The research was entirely confidential. The results of the study provide a summary of common characteristics and do not reveal any identifying characteristics or information of individual students. This research was deemed non-reviewable by the IRB because it does not involve human subjects, since de-identified data sets cannot be traced back to the original subjects.

Limitations

There exist some limitations to this study. The first is the relatively small sample size (n = 119). The sample size is such that it is difficult to generalize the findings to a larger population of students, particularly for a group of ELs that may be more linguistically homogeneous. The study does however aim to contribute to the body of educational research on the topic of standardized assessment practices for ELs, and the relation of second language proficiency to achievement on standardized content-area assessments.

Another limitation is the pre-existing differences among the EL student population used in the study. Each child has a unique educational background and personal experience. The amount of time spent in formal schooling in their primary language, length of residence in the United States, instructional strategies and effectiveness of the classroom teacher, language and academic support available outside of school, and various other factors may pose some threat to the validity of this study. Each student enters this study in a different place on their educational journey.

Similarly, the study has limitations with respect to the generalizability of the racial and ethnic identities of the student population. Racial/ethnic categories are broadly defined and encompass many different and unique identities. Accordingly, the ability to generalize this study along racial/ethnic findings is limited.

Finally, the use of standardized assessments itself to measure language proficiency and knowledge more generally is an inherent limitation. As this study and a wealth of supporting research detail, language is complex, dynamic, and socially constructed. To attempt to define one's linguistic capabilities using a standardized measure will surely miss the mark on the range of possibilities for any individual. As Baker and Wright (2017) note:

> The sub-components of language proficiency are not easily definable or measurable. Apart from language skills, there are the qualitative aspects of language that are not simply reducible for testing (e.g. the emotive, status and poetic functions of languages). There is growing recognition that a single test cannot provide an accurate measure of a bilingual's proficiency. (p. 23)

The same can be said for the use of standardized assessments in assessing human knowledge and expected learning. Thus, and perhaps ironically, the use of standardized assessment data to analyze the impact of such data is in itself a limitation to the study.

Conclusion

This study discusses the issues related to the challenges of the inclusion of ELs within a system of high-stakes accountability. The study seeks to address the overarching issue of the relationship between an EL's second language proficiency and their performance on federally mandated standardized content-area assessments. Currently the State of Illinois is at a crossroads

as it makes decisions concerning the format and structure of its accountability assessments beginning in the 2021-2022 school year. It is hoped that this study will help develop a more clear empirical understanding of the relationship between English language proficiency and grade level content-area achievement as measured by standardized assessments that can be used to inform decision-making for ELs at the federal, state, district, school, and classroom level. The equity in assessment and educational experience for ELs is critical to our work in schools of providing all students with the knowledge, understandings, strategies, and confidence to achieve success in a self-determined future.

CHAPTER FOUR. FINDINGS AND ANALYSIS OF THE DATA.

Introduction

This chapter provides an overview of the data and analysis of the results relating to each of the research questions and underlying hypotheses. The purpose of this quasi-experimental correlational design quantitative study was to examine the relationship between ELs' second language proficiency and achievement on federally mandated standardized content-area assessments. A more clearly developed empirical understanding of the relationship between English language proficiency and grade level content-area achievement, as measured by performance-based standardized tests, will help to inform both legislators and educators as to assessment practices and policies that are most equitable and valid for ELs and their school communities.

As discussed in the previous chapter, this study utilized three different instruments for data analysis. The ACCESS for ELLs 2.0 assessment was the instrument used to measure an EL's level of language proficiency in English. The PARCC assessment was the instrument utilized to measure grade level content-area proficiency in ELA and mathematics. The final instrument used for this study was student registration information from the 2017-2018 school year. This included the demographic information that families reported as part of their initial paperwork upon registering within the district at the beginning of the school year. The relationship between the ACCESS and PARCC scores for a small, yet diverse, group of third through fifth grade ELs was investigated and analyzed in this quasi-experimental correlational design quantitative study.

The overall research question guiding this study was: How does an ELs' second language proficiency influence his/her achievement on federally mandated standardized content-area assessments? The four research questions and seven corresponding hypotheses are utilized as a framework for presenting the findings of this study. The data collection procedures and results of the statistical tests applied for each question are reviewed, and are followed by an analysis of the data for each research question. The chapter concludes with a summary of the findings.

Research Question One

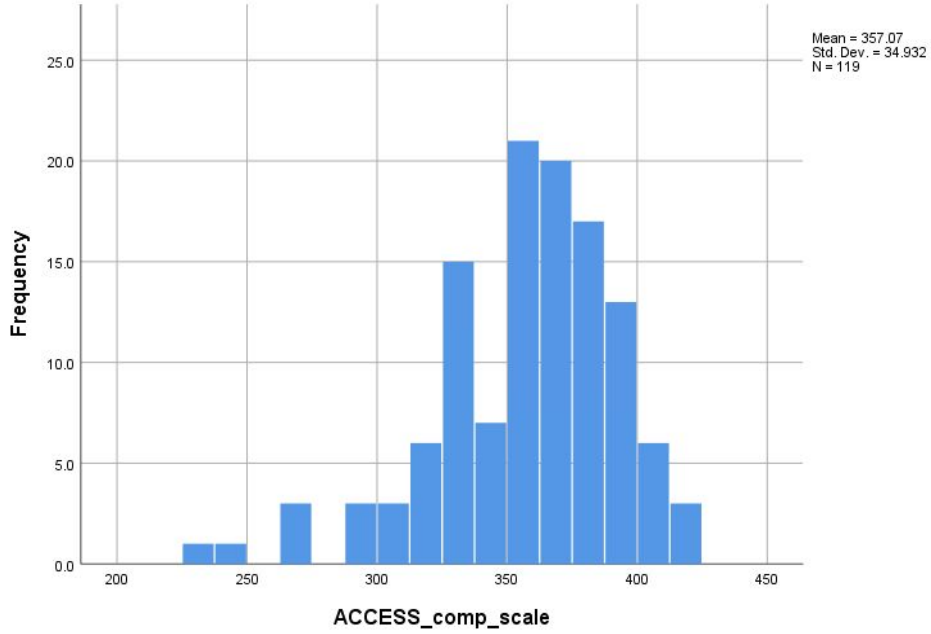
The first research question guiding this study was: what is the relationship between English language proficiency and achievement on grade level content-area standardized assessments in ELA and mathematics? The initial hypothesis for this question was: the more advanced the student's level of language proficiency in English, the higher their reading achievement.

Linear correlation was the statistical test utilized to answer the first research question and test the underlying hypotheses. For the first hypothesis, composite ACCESS scale scores were correlated with PARCC reading scale scores to determine the strength and direction of the relationship. The primary assumptions for correlation as a statistical procedure are: a randomly chosen sample, variables that are independent of one another, variables that are normally distributed, equal variances, and a linear relationship (Abbott, 2011). In this study, all assumptions with the exception of a randomly chosen sample were met. The data sets were not linked and thus independent of one another. Both sets of data represent scaled scores and were interval level. *Table 5* shows the descriptive statistics for the study variables; the variables appear normal from the results.

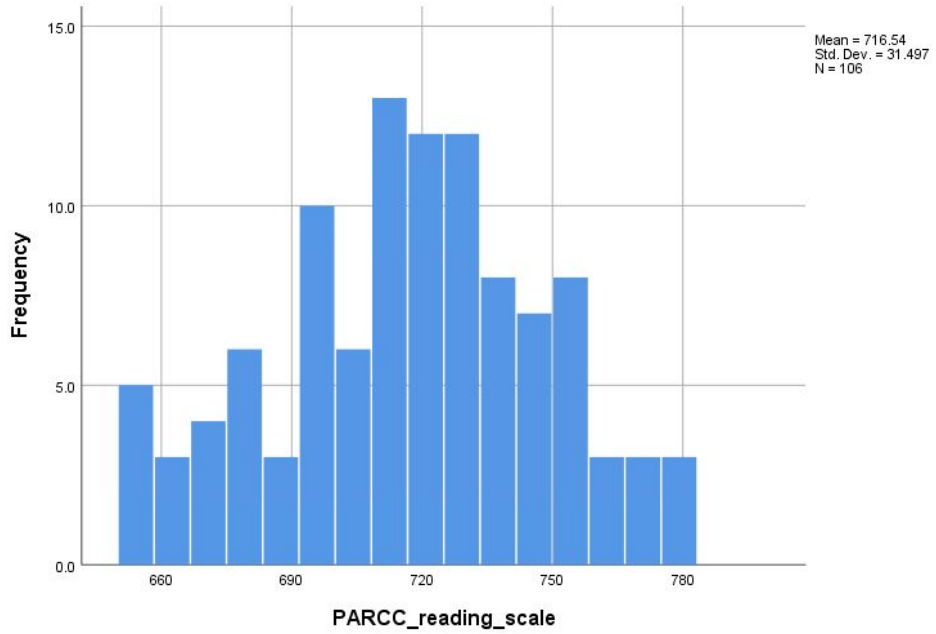
Table 5									
Descriptive Statistics									
	N	Mean	Std. Deviation	Variance	Range	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
ACCESS composite scale score	119	357.070	34.932	1220.216	229 to 420	-1.066	0.222	1.720	0.440
PARCC ELA scale score	106	716.540	31.497	992.080	650 to 780	-0.199	0.235	-0.482	0.465
Valid N	106								

Table 5. Descriptive Statistics for Correlation of ACCESS Composite Scale Scores and PARCC ELA Scale Scores.

Graph 1 shows the histogram for ACCESS composite scale scores. *Graph 2* depicts the histogram for PARCC ELA scores. As can be seen on the pair of graphs, both variables appear normally distributed. The variances can be considered equal as no graph is markedly skewed.

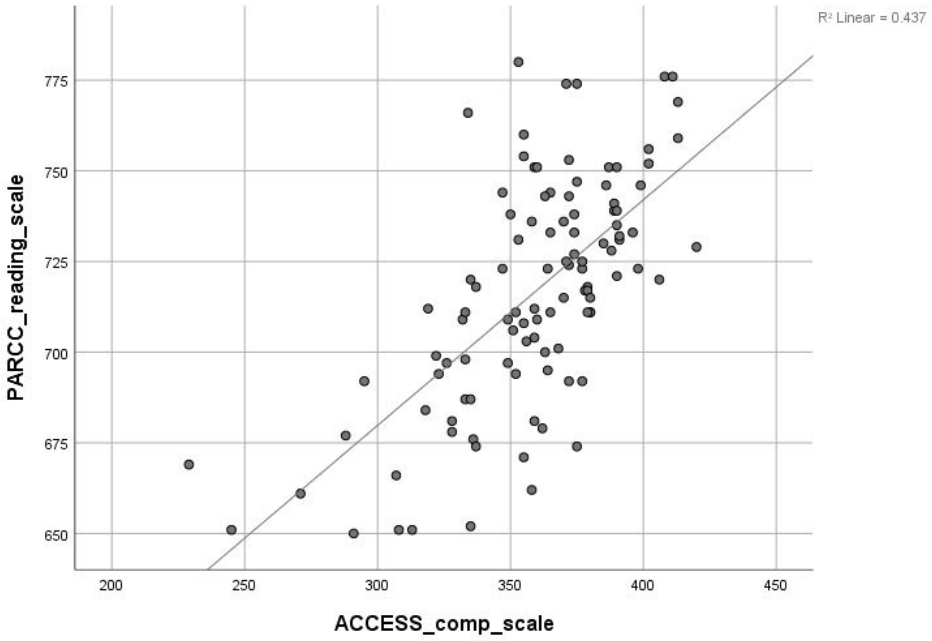


Graph 1. Histogram of ACCESS Composite Scale Scores.



Graph 2. Histogram of PARCC ELA Scale Scores.

Graph 3, a scattergram depicting PARCC reading scale scores and ACCESS composite scale scores, shows a linear relationship between the two study variables. This visual pattern indicates a positive correlation. As a student's ACCESS composite scale score increases in value the PARCC reading scale score also increases. The pattern of dots is fairly evenly distributed around the line of best fit, indicating a relatively strong and statistically significant relationship between the two variables. The overall shape of the dots is an upward sloping straight line.



Graph 3. Scattergram of PARCC Reading Scale Scores and ACCESS Composite Scale Scores.

Even with the slight violation of the first assumption regarding randomly chosen samples, agreement with all of the other assumptions enables linear correlation to still be utilized as a statistical test to answer the first research question and test its two underlying hypotheses. Linear correlation is a robust test, and can provide meaningful results even with slight violations of its assumptions (Abbott, 2011).

Table 6 presents the data for the linear correlation of ACCESS composite scale scores and PARCC ELA scale scores. The following analysis employed three methods to determine the strength of the relationship between a student's English language proficiency and achievement on grade level content-area standardized assessments in ELA in response to the first research question and its first hypothesis.

Table 6			
Correlations			
		ACCESS composite scale score	PARCC ELA scale score
ACCESS composite scale score	Pearson Correlation	1	0.661**
	Sig. (2-tailed)		0.000
	N	119	106
PARCC ELA scale score	Pearson Correlation	0.661**	
	Sig. (2-tailed)	0.000	
	N	106	106

**. Correlation is significant at the 0.01 level (2-tailed).

Table 6. Correlations Table for ACCESS Composite Scale Scores and PARCC ELA Scale Scores.

The Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis states that the relation between the ACCESS composite proficiency level and PARCC ELA scale score is 0. The Alternative Hypothesis states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table (Appendix A) specifies that the two-tailed 0.05 critical value for *r* at 104 degrees of freedom is rdf(.05) = 0.195. Hence, values greater than 0.195 are considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.661$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.661 is statistically significant.

The second method applied for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.661 was transformed using the formula from Cohen (1988) into a *t* ratio of 9.01. The *T* table of values has a critical value of t 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.661) is significant because it resulted in a *t* score (9.01) that exceeded the critical *T* value (1.980).

The final method for assessing the strength of the relationship between language proficiency and reading achievement was to examine the effect size. With correlation, the strength of the relationship between two variables can be assessed using the coefficient of determination or r^2. According to Cohen (1988) a large effect size has an r^2 value of 0.25 (r = 0.50) or greater. Using this measure our r^2 value of 0.437 represents a large effect size, indicating a strong and statistically significant relationship between a student's composite ACCESS score and their PARCC reading scale score.

Through utilizing multiple methods to assess the strength of the relationship, it is evident that a student's composite ACCESS proficiency score positively correlates to their PARCC reading scale score. With a Pearson Coefficient, r, of 0.661 the strength of the relationship as measured by r^2 is 0.437. As depicted in *Figure 16,* approximately 44% of the variance in scores is explained by these two variables.

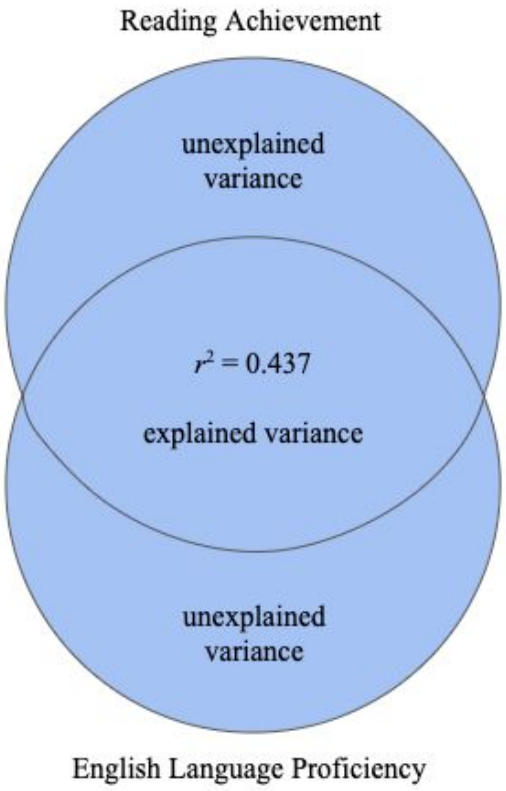


Figure 16. The Effect Size of Correlation - Reading Achievement and English Language Proficiency.

Nearly half of a student's PARCC reading score can be explained by their composite ACCESS score. This empirical understanding reflects the role of English language proficiency as a key indicator of academic achievement for ELs on the PARCC standardized assessment for ELA/literacy.

The Chi-Square test of independence was used to detect patterns that may indicate relatedness between the two variables. The assumptions for the Chi-Square test of independence are: variables that are independent of one another and having no expected cell frequencies that are less than five (Abbott, 2011). In this study, the cell frequency assumption was violated due to the smaller sample size. The number of cells with a count of less than five (41.7%) is noted

on the contingency table. Even with the violation, this statistical procedure can still provide meaningful results for the study (Abbott, 2011). *Table 7* shows the results of the Chi-Square test of independence between the English proficiency levels of EL students and grade level reading proficiency. *Table 7* shows that the hypothesis of independence between English language proficiency and reading proficiency is rejected - the two-sided *p*-value is 0.000.

Table 7

Reading achievement of EL students and English language proficiency

	Reading Proficiency					
English language proficiency	Level 1: did not yet meet expectations	Level 2: partially met expectations	Level 3: approached expectations	Level 4: met expectations	Level 5: exceeded expectations	Total percentage
	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)
	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)
Low	100 (6)	0	0	0	0	100 (6)
	19.4 (6)	0	0	0	0	5.7 (6)
Medium	33.3 (25)	36 (27)	21.3 (16)	9.3 (7)	0	100 (75)
	80.6 (25)	87.1 (27)	59.3 (16)	41.2 (7)	0	70.8 (75)
High	0	16 (4)	44 (11)	40 (10)	0	100 (25)
	0	12.9 (4)	40.7 (11)	58.5 (10)	0	23.6 (25)
Total (N= 106)	29.2 (31)	29.2 (31)	25.5 (27)	16 (17)	0	100 (106)
	100 (31)	100 (31)	100 (27)	100 (17)	0	100 (106)

Pearson Chi-Square (df = 6) = 39.852; two-sided p-value = .000

*** 5 cells (41.7%) have expected count less than 5. The minimum expected count is .96

Table 7. Contingency Table and Results of Chi-Square Test of Independence for English Proficiency Levels and Grade Level Reading Proficiency Levels.

As illustrated above, students in the low English language proficiency category all scored at a level 1 (did not yet meet expectations). Also demonstrated in *Table 7*, is the increase of reading proficiency as students acquire more English language proficiency; 66.7% of medium English language proficiency students scored at a level 2 (partially met expectations) or above, while 40% of high English language proficiency students scored a level 4 (met expectations) for grade level reading proficiency. The Chi-Square value was significant at 39.85. Given our calculated value (39.852) exceeded the critical value (15.507), it can be concluded that there is a statistically significant difference amongst the reading achievement of EL students based on their level of English proficiency at the 0.05 level. Thus, English language proficiency is clearly a

relational factor of reading achievement for EL students on the PARCC ELA assessment, a federally mandated performance-based standardized assessment.

Based upon linear correlation and the Chi-Square test of independence it is evident that the first hypothesis is supported by the data. The evidence from this student population indicates that the more advanced a student's level of language proficiency in English, the higher their reading achievement.

The second hypothesis for the first research question was: the more advanced the student's level of language proficiency in English, the higher their mathematics achievement. Linear correlation was again the statistical test utilized to answer the first research question and test its hypotheses. For the second hypothesis, composite ACCESS scale scores were correlated with PARCC math scale scores to determine the strength and direction of the relationship. As with the first hypothesis, all assumptions with the exception of a randomly chosen sample were met. The data sets were not linked and thus independent of one another. Both data sets represent scaled scores and were interval level. *Table 8* shows the descriptive statistics for the study variables; the variables appear to be normally distributed from the results.

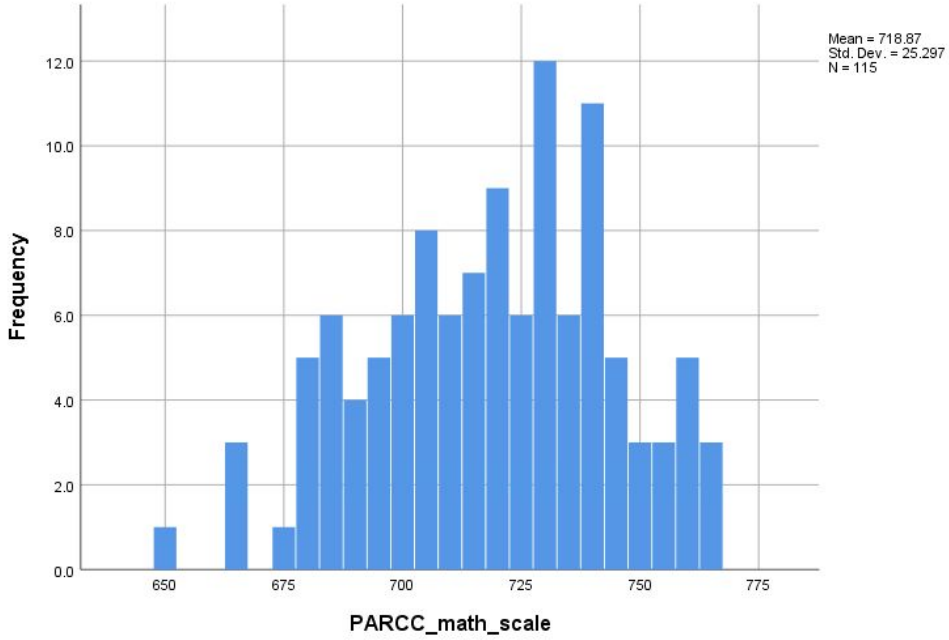
Table 8

Descriptive Statistics

	N	Mean	Std. Deviation	Variance	Range	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
ACCESS composite scale score	119	357.070	34.932	1220.216	229 to 420	-1.066	0.222	1.720	0.440
PARCC math scale score	115	718.870	25.297	639.939	650 to 766	-0.234	0.226	-0.482	0.447
Valid N	115								

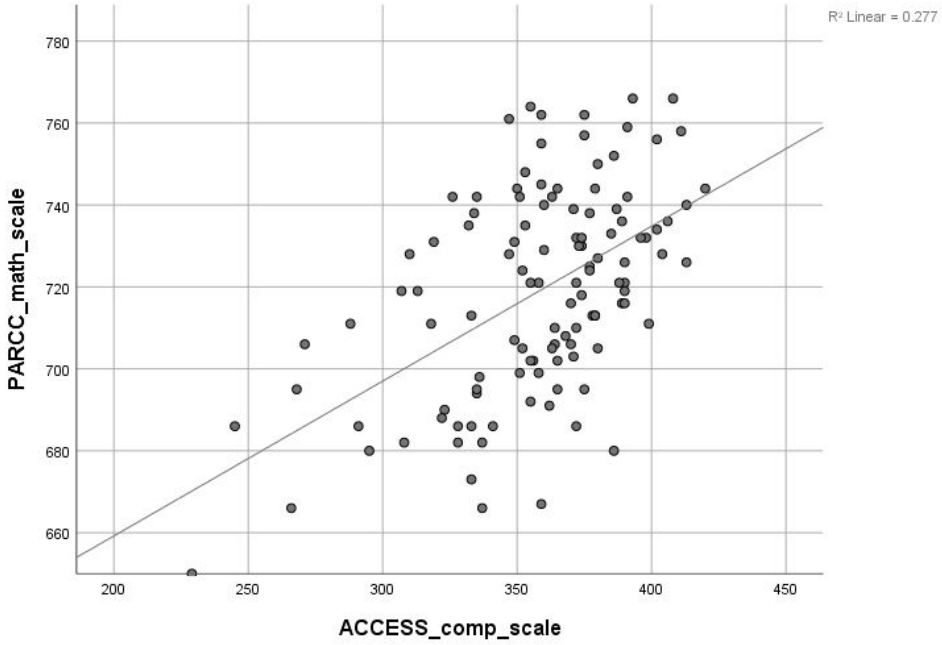
Table 8. Descriptive Statistics for Correlation of ACCESS Composite Scale Score and PARCC Math Scale Score.

Graph 1 shows the histogram for ACCESS composite scale scores. *Graph 4* depicts the histogram for PARCC math scores. As can be seen on the pair of graphs, both variables appear normally distributed. The variances can be considered equal as no graph is markedly skewed.



Graph 4. Histogram of PARCC Math Scale Scores.

Graph 5, a scattergram depicting PARCC math scale scores and ACCESS composite scale scores, shows a linear relationship between the two study variables. This visual pattern indicates a positive correlation. As a student's ACCESS composite scale score increases in value, the PARCC math scale score also increases. The pattern of dots is fairly evenly distributed around the line of best fit, indicating a relatively strong and statistically significant relationship between our two variables. The overall shape of the dots is an upward sloping straight line.



Graph 5. Scattergram of PARCC Math Scale Scores and ACCESS Composite Scale Scores.

Table 9 presents the data for the linear correlation of ACCESS composite scale scores and PARCC math scale scores. The following analysis employed three methods to determine the strength of the relationship between a student's English language proficiency and achievement on grade level content-area standardized assessments in mathematics in response to the first research question and its second hypothesis.

Table 9

Correlations

		ACCESS composite scale score	PARCC math scale score
ACCESS composite scale score	Pearson Correlation	1	0.526**
	Sig. (2-tailed)		0.000
	N	119	115
PARCC math scale score	Pearson Correlation	0.526**	
	Sig. (2-tailed)	0.000	
	N	115	115

**. Correlation is significant at the 0.01 level (2-tailed).

Table 9. Correlations Table for ACCESS Composite Scale Scores and PARCC Math Scale Scores.

The Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis states that the relation between the ACCESS composite proficiency level and PARCC mathematics scale score is 0. The Alternative Hypothesis states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table (Appendix A) specifies that the two-tailed 0.05 critical value for *r* at 113 degrees of freedom is rdf(.05) = 0.195. Hence, values greater than 0.195 would be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.526$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.526 is statistically significant.

The second method applied for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.526 was transformed using the formula from Cohen (1988) into a *t* ratio of 10.71. The *T* table of values has a critical value of *t* 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.526) is significant because it resulted in a *t* score (10.71) that exceeded the critical *T* value (1.980).

The final method for assessing the strengths of the relationship between language proficiency and math achievement was to examine the effect size. According to Cohen (1988) a large effect size has an r^2 value of 0.25 (r = 0.50) or greater. Using this measure our r^2 value of 0.277 represents a large effect size, indicating a strong and statistically significant relationship between a student's composite ACCESS score and their PARCC math scale score.

Utilizing multiple methods to assess the strength of the relationship, it is evident that a student's ACCESS composite proficiency score positively correlates to their PARCC math scale score. With a Pearson Coefficient, *r*, of 0.526 the strength of the relationship as measured by r^2 is 0.277. As shown in *Figure 17*, approximately 27% of the variance in scores is explained by these two variables.

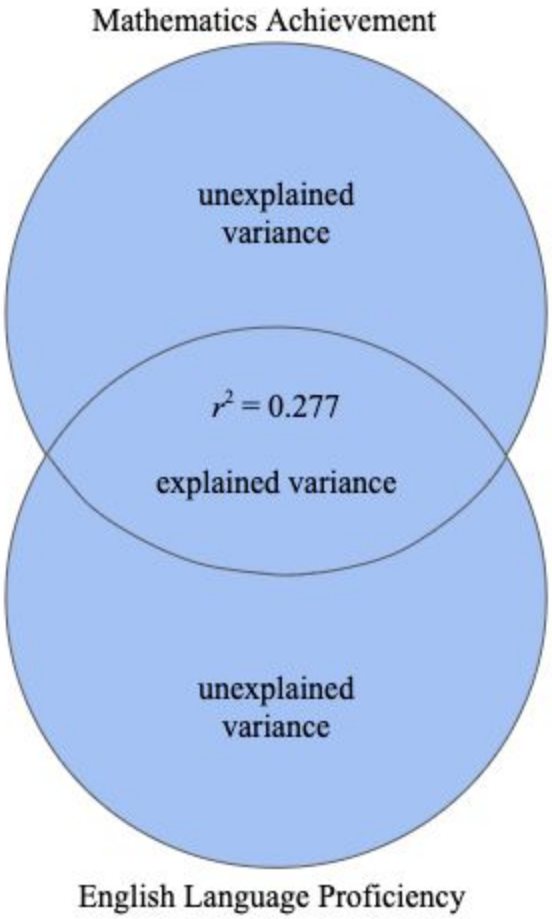


Figure 17. The Effect Size of Correlation - Reading Achievement and English Language Proficiency.

Just over a quarter of a student's PARCC math score can be explained by their composite ACCESS score. Though a less significant correlation than was seen for reading, this empirical

understanding reflects the role of English language proficiency as a key indicator of academic achievement for ELs on the PARCC standardized assessment for mathematics.

Similar to testing the first hypothesis, the Chi-Square test of independence was used to detect patterns that may indicate relatedness between the two variables. Again the cell frequency assumption is violated due to the smaller sample size. The number of cells with a count of less than five (41.7%) is noted on the contingency table. Even with the violation, this statistical procedure can still provide meaningful results for the study (Abbott, 2011). *Table 10* shows the results of the Chi-Square test of independence between the English proficiency levels of EL students and grade level mathematics proficiency. *Table 10* shows that the hypothesis of independence between English language proficiency and mathematics proficiency is rejected - the two-sided *p*-value is 0.000.

Table 10

Mathematics achievement of EL students and English language proficiency

	Mathematics Proficiency					
English language proficiency	Level 1: did not yet meet expectations	Level 2: partially met expectations	Level 3: approached expectations	Level 4: met expectations	Level 5: exceeded expectations	Total percentage
	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)
	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)
Low	75 (6)	25 (2)	0	0	0	100 (8)
	21.4 (6)	5.7 (2)	0	0	0	7.0 (8)
Medium	27.5 (22)	35 (28)	31.3 (25)	6.3 (5)	0	100 (80)
	78.6 (22)	80.0 (28)	64.1 (25)	38.5 (5)	0	69.5 (80)
High	0	18.5 (5)	51.9 (14)	29.6 (8)	0	100 (27)
	0	14.3 (5)	35.9 (14)	61.5 (8)	0	23.5 (27)
Total (N= 115)	24.3 (28)	30.4 (35)	34.0 (39)	11.3 (13)	0	100 (115)
	100 (28)	100 (35)	100 (39)	100 (13)	0	100 (115)

Pearson Chi-Square (df = 6) = 33.391; two-sided p-value = .000

*** 5 cells (41.7%) have expected count less than 5. The minimum expected count is .90

Table 10. Contingency Table and Results of Chi-Square Test of Independence for English Proficiency Levels and Grade Level Mathematics Proficiency Levels.

As illustrated above, students in the low English language proficiency category all scored at a level 1 (did not yet meet expectations) or level 2 (partially met expectations). Also demonstrated in *Table 10*, is the increase of mathematics proficiency as students acquire more English

language proficiency; 72.5% of medium English language proficiency students scored at a level 2 (partially met expectations) or above, while 29.6% of high English language proficiency students scored a level 4 (met expectations) for grade level math proficiency. The Chi-Square value is significant at 33.391. As the calculated value (33.391) exceeded the critical value (15.507), it can be concluded that there is a statistically significant difference amongst the mathematics achievement of EL students based on their level of English proficiency at the 0.05 level. Thus, English language proficiency is clearly a relational factor of math achievement for EL students on the PARCC mathematics assessment, a federally mandated performance-based standardized assessment.

Based upon linear correlation and the Chi-Square test of independence it is evident that the second hypothesis is supported by the data. The evidence from this student population indicates that the more advanced a student's level of language proficiency in English, the higher their mathematics achievement.

Research Question Two

The second research question guiding this study was: to what extent does English language proficiency in reading, writing, speaking, and listening influence achievement on grade level content-area standardized assessments in ELA and mathematics? The initial hypothesis for this question was: a student's English language proficiency in reading and writing will impact their reading achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.

Linear correlation and multiple linear regression were the statistical tests utilized to answer the second research question and test its underlying hypotheses. For this hypothesis, the

PARCC scale scores for ELA were the dependent variable, with the scale scores for each of the four ACCESS assessed language domains - speaking, listening, reading, and writing - as the independent variables. The primary assumptions for both statistical tests are similar: interval level variables, variables that are normally distributed, equal variances, variables that are independent of one another, and a linear relationship (Abbott, 2011). With respect to correlation, all assumptions were met with the exception of a randomly selected sample. Again, correlation is a robust test and can still provide meaningful results even with slight violations of its assumptions (Abbott, 2011). For multiple linear regression all of the assumptions were met. The variables are normally distributed and interval level. *Table11* shows the descriptive statistics for the study variables; the variables appear normal from the results.

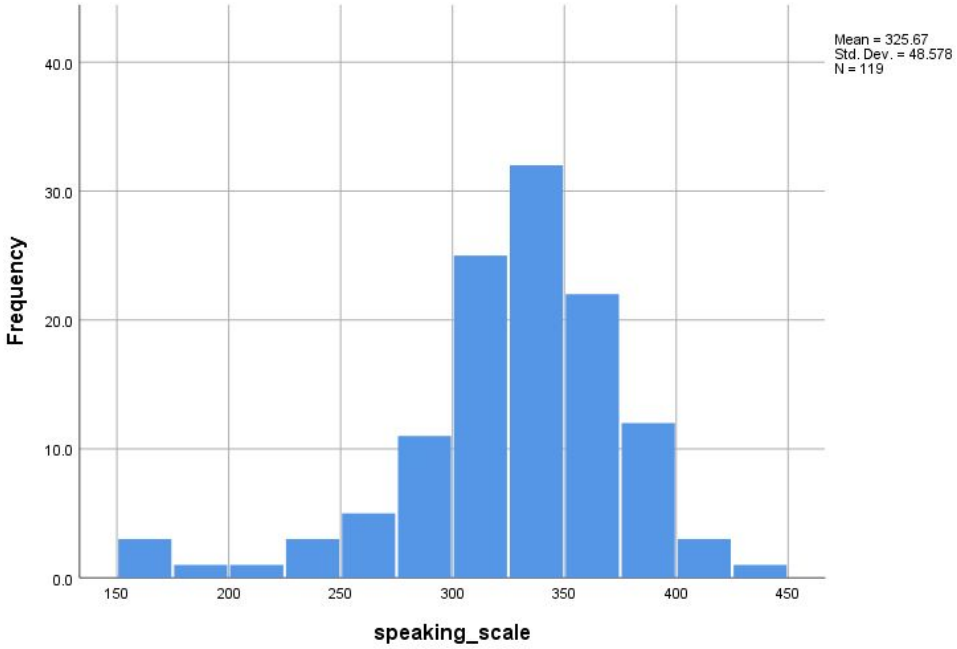
Table 11

Descriptive Statistics

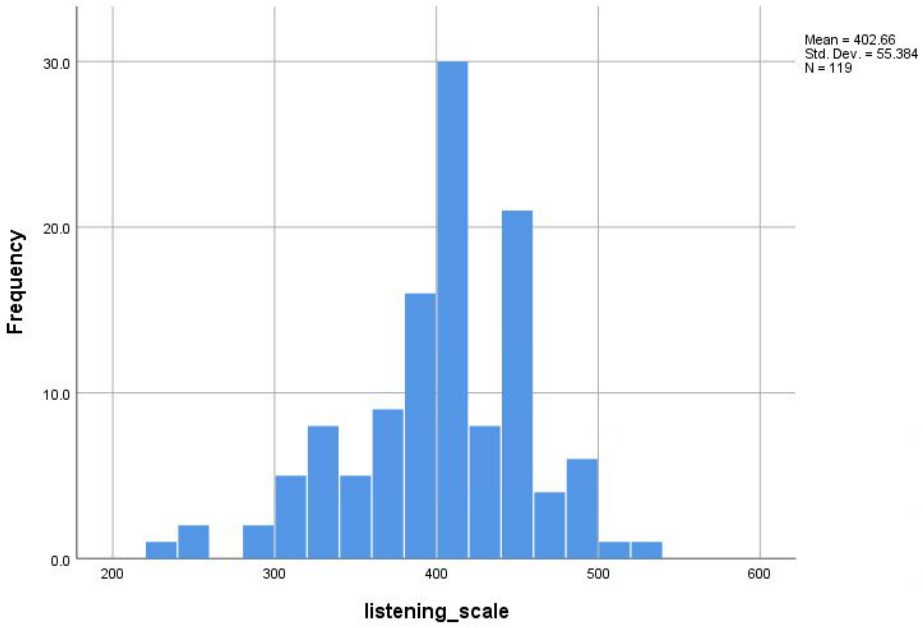
	N	Mean	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
PARCC ELA scale score	106	716.540	31.497	992.080	-0.199	0.235	-0.482	0.465
ACCESS listening scale score	119	402.660	55.384	3067.380	-0.750	0.222	0.806	0.440
ACCESS speaking scale score	119	325.670	48.578	2359.798	-1.312	0.222	2.921	0.440
ACCESS reading scale score	119	357.270	35.546	1263.520	-0.544	0.222	-0.440	0.440
ACCESS writing scale score	119	350.690	37.646	1417.233	-1.764	0.222	8.348	0.440

Table 11. Descriptive Statistics for Multiple Linear Regression of PARCC ELA Scale Scores and ACCESS Language Domains.

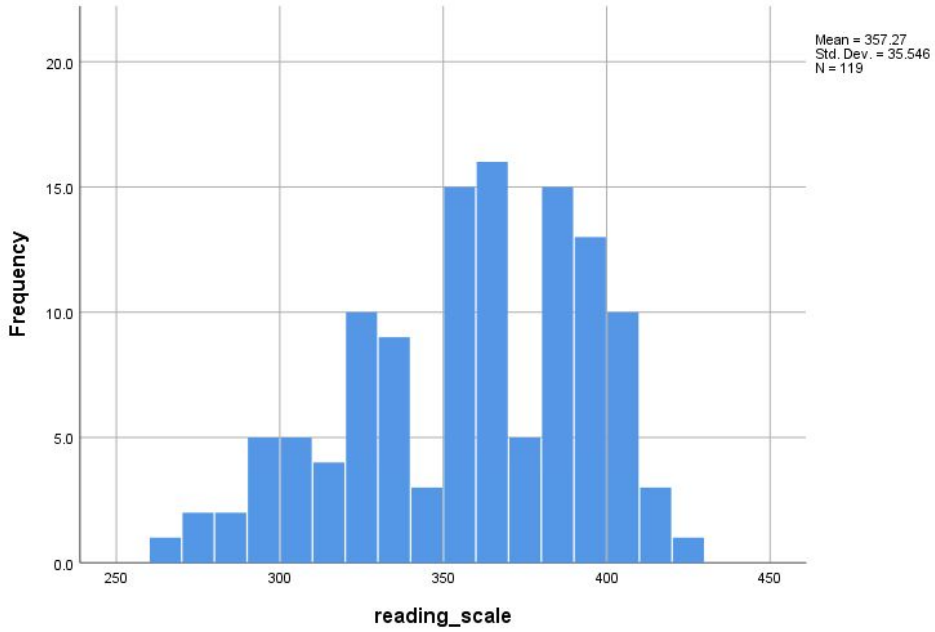
Graph 2 depicts the histogram for PARCC ELA scores. *Graphs 6, 7, 8,* and *9* show the histograms for each of the four ACCESS assessed language domains - speaking, listening, reading, and writing. The variables appear normally distributed on each of the graphs. The variances can be considered equal as no graph is markedly skewed.



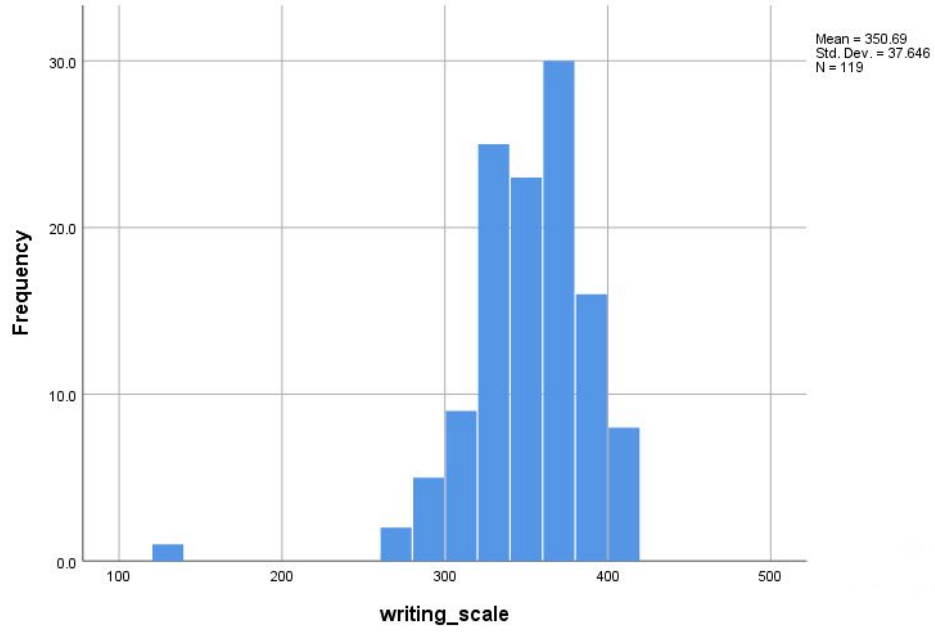
Graph 6. Histogram of ACCESS Speaking Scale Scores.



Graph 7. Histogram of ACCESS Listening Scale Scores.

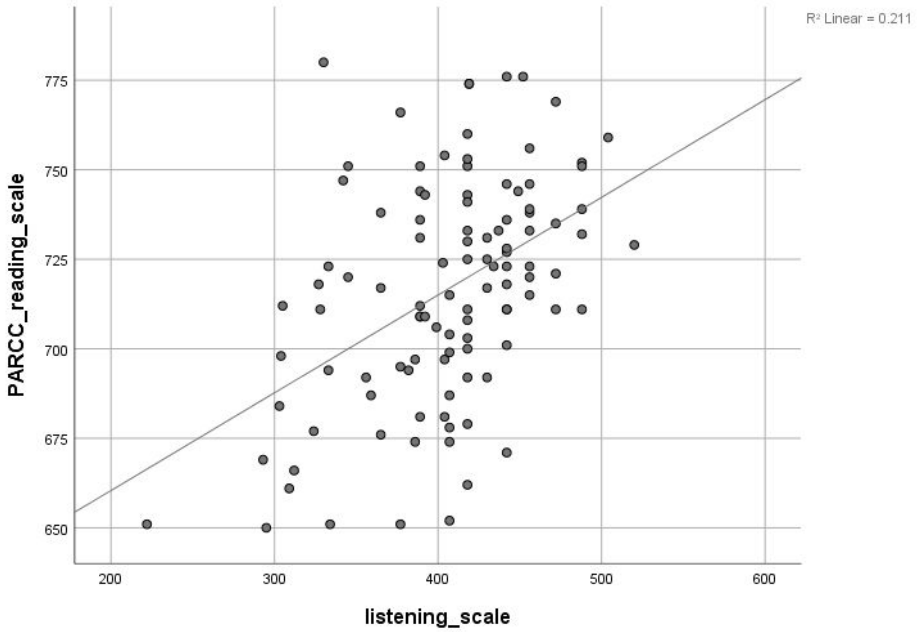


Graph 8. Histogram of ACCESS Reading Scale Scores.

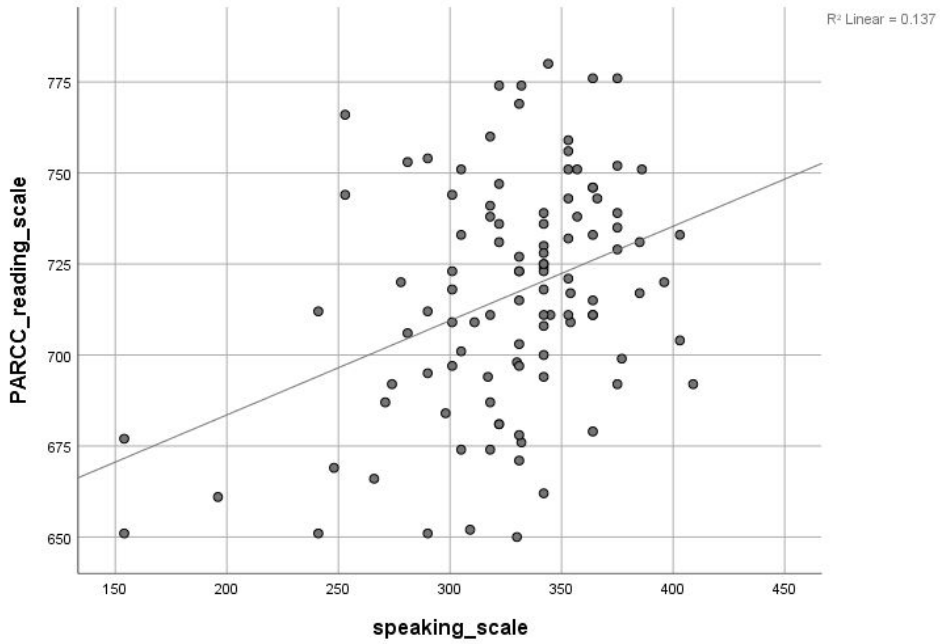


Graph 9. Histogram of ACCESS Writing Scale Scores.

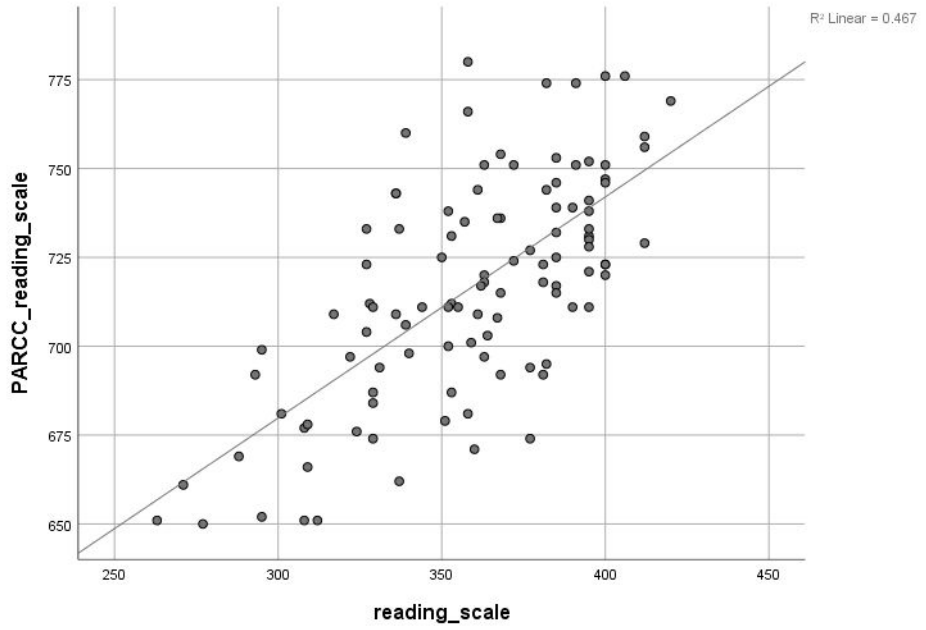
Graphs 10, 11, 12, and *13* are scattergrams depicting PARCC ELA scale scores and each of the four ACCESS language domains - listening, speaking, reading, and writing. Each graph illustrates, to varying degrees, a linear relationship between the variables. *Graphs 12* and *13* show the strongest correlation, exhibiting the relationship between PARCC ELA scale scores and ACCESS reading and writing scale scores. The pattern of dots are fairly evenly distributed around the line of best fit, indicating a relatively strong and statistically significant relationship between these variables. The overall shape of the dots is an upward sloping straight line for both graphs. These visual patterns indicate a positive correlation. As a student's ACCESS reading and writing scale score increases in value their PARCC ELA scale score also increases. *Graphs 10* and *11* show less significant, but still visible correlations between PARCC ELA scale scores and ACCESS listening and speaking scale scores. The overall cluster of dots is a bit broader, but still slopes upward reflecting a relationship between the variables.



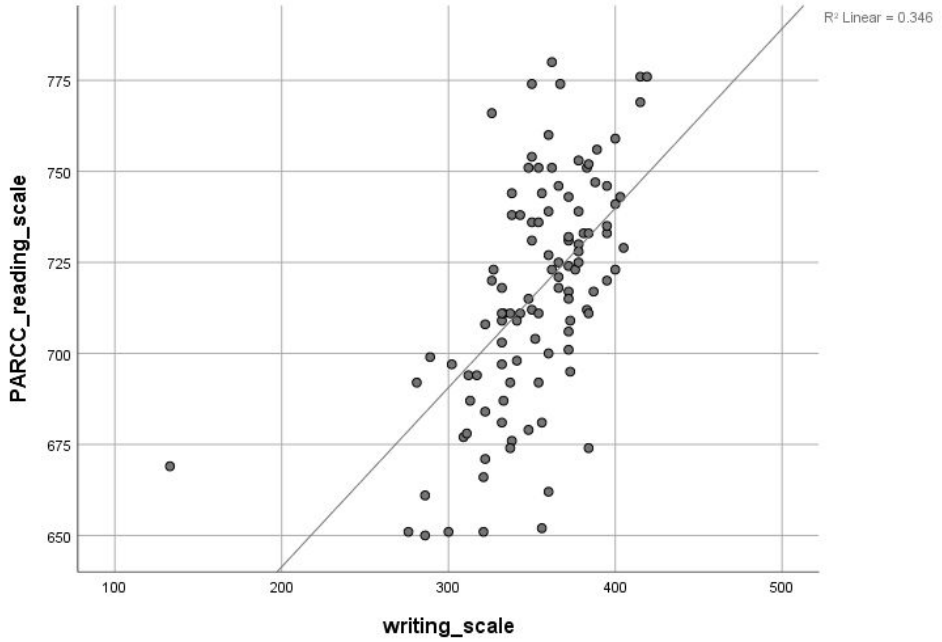
Graph 10. Scattergram of PARCC ELA Scale Scores and ACCESS Listening Scale Scores.



Graph 11. Scattergram of PARCC ELA Scale Scores and ACCESS Speaking Scale Scores.



Graph 12. Scattergram of PARCC ELA Scale Scores and ACCESS Reading Scale Scores.



Graph 13. Scattergram of PARCC ELA Scale Scores and ACCESS Writing Scale Scores.

Table 12 presents the data for the correlation of each of the four ACCESS assessed language domains and PARCC ELA scale scores. The following analysis employed three methods to determine the strength of the relationship between a student's proficiency in each of the four ACCESS assessed language domains - listening, speaking, reading, writing - and grade level content-area standardized assessments in ELA in response to the second research question and its hypothesis.

Table 12

Pearson Correlation

	PARCC ELA scale score	ACCESS listening scale score	ACCESS speaking scale score	ACCESS reading scale score	ACCESS writing scale score
PARCC ELA scale score	1.000	0.460	0.370	0.683	0.588
ACCESS listening scale score	0.460	1.000	0.597	0.703	0.563
ACCESS speaking scale score	0.370	0.597	1.000	0.483	0.490
ACCESS reading scale score	0.683	0.703	0.483	1.000	0.662
ACCESS writing scale score	0.588	0.563	0.490	0.662	1.000

Table 12. Pearson Correlation for PARCC ELA Scale Scores and ACCESS Language Domains.

ACCESS listening scale scores were considered first. The Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis (H0): p = 0, states that the relation between ACCESS listening scale scores and PARCC ELA scale scores is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 104 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would therefore be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.460$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.460 is statistically significant.

The second method for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.460 was transformed using the formula from Cohen (1988) to a *t* ratio of 6.38. The *T* table of values has a critical value of *t* 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.460) is significant because it resulted in a *t* score (6.38) that exceeded the critical *T* value (1.980).

The final method for assessing the strength of the relationship between the listening domain of language proficiency and reading achievement was to examine the effect size. With correlation, the strength of the relationship between two variables can be assessed using the coefficient of determination or r^2. According to Cohen (1988) a medium effect size has an r^2 value of 0.09 ($r = 0.30$) or greater, while a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.21 represents a medium effect size, indicating a relatively strong relationship between ACCESS listening scores and PARCC ELA scale scores.

ACCESS speaking scale scores were examined next. The Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis (H0): p = 0, states that the relation between ACCESS speaking scale scores and PARCC reading scale scores is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 104 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would therefore be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.370$) exceeds the critical value of exclusion (0.195). A significant

correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.370 is statistically significant.

The second method for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.370 was transformed using the formula from Cohen (1988) to a *t* ratio of 4.75. The *T* table of values has a critical value of *t* 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.370) is significant because it resulted in a *t* score (4.75) that exceeded the critical *T* value (1.980).

The final method for assessing the strengths of the relationship between the speaking domain of language proficiency and reading achievement was to examine the effect size. With correlation, the strength of the relationship between two variables can be assessed using the coefficient of determination or r^2. According to Cohen (1988) a medium effect size has an r^2 value of 0.09 ($r = 0.30$) or greater, while a large effect size has an *r* squared value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.14 represents a medium effect size, indicating a relatively strong relationship between ACCESS speaking scores and PARCC ELA scale scores.

With respect to ACCESS reading scale scores, the Hypothesis Test for Pearson's *r* will be utilized first. The Null Hypothesis (H0): p = 0, states that the relation between the ACCESS reading scale scores and PARCC reading scale scores is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 104 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would

therefore be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.683$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.683 is statistically significant.

The second method of judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.683 was transformed using the formula from Cohen (1988) to a *t* ratio of 12.37. The *T* table of values has a critical value of t 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.683) is significant because it resulted in a *t* score (12.37) that exceeded the critical *T* value (1.980).

The final method for assessing the strengths of the relationship between the reading domain of language proficiency and reading achievement was to examine the effect size. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.470 represents a large effect size, indicating a strong and statistically significant relationship between ACCESS reading scores and PARCC reading scale scores.

For ACCESS writing scale scores, the Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis (H0): p = 0, states that the relation between ACCESS writing scale scores and PARCC reading scale score is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 104 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would therefore be considered

statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.588$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.588 is statistically significant.

The second method for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.588 was transformed using the formula from Cohen (1988) to a *t* ratio of 9.34. The *T* table of values has a critical value of t 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.588) is significant because it resulted in a *t* score (9.34) that exceeded the critical *T* value (1.980).

The final method for assessing the strengths of the relationship between the written domain of language proficiency and reading achievement was to examine the effect size. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.35 represents a large effect size, indicating a strong relationship between ACCESS writing scores and PARCC reading scale scores.

Table 13 displays the summary data for the linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.485, equates to approximately 49% of the variance in PARCC reading scores explained by the four language proficiencies of the ACCESS assessment - listening, speaking, reading, and writing.

Table 13									
Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	0.710*	0.504	0.485	22.611	0.504	25.685	4	101	0.000
* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale									

Table 13. Multiple Linear Regression Model Summary.

Table 14 displays the ANOVA summary table for the multiple linear regression model. The *F* test shows a value of 25.685 with a significance of 0.000, less than the 0.05 threshold to be considered statistically significant.

Table 14						
*ANOVA**						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	52529.295	4	13132.324	25.685	.000**
	Residual	51639.054	101	511.278		
	Total	104168.349	105			
* Dependent Variable: PARCC_reading_scale						
** Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale						

Table 14. ANOVA Summary.

Table 15 depicts the coefficient outputs for the regression model. When examining the coefficient data of the multiple linear regression model, the impact of each ACCESS language proficiency on the PARCC reading scale scores can be seen. The ACCESS reading scale score, with a *p*-value of 0.000, is statistically significant. The ACCESS writing scale score is also significant with a *p*-value of 0.011. Both the ACCESS listening and speaking scale score are not significant with *p*-values of 0.335 and 0.731 respectively.

Table 15

Coefficients

Model		Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B	
							Lower Bound	Upper Bound
1	(constant)	472.304	24.556		19.234	0.000	423.592	521.015
	ACCESS listening scale score	-0.062	0.064	-0.105	-0.966	0.336	-0.190	0.066
	ACCESS speaking scale score	0.022	0.063	0.031	0.345	0.731	-0.103	0.147
	ACCESS reading scale score	0.524	0.101	0.575	5.211	0.000	0.324	0.723
	ACCESS writing scale score	0.211	0.081	0.252	2.591	0.011	0.049	0.372

* Dependent Variables: PARCC_reading_scale

Table 15. Coefficient Outputs.

Accordingly, utilizing the unstandardized *B* coefficients it is predicted that a one point increase on the ACCESS reading scale score equals a 0.524 point increase of the PARCC reading scale score. A one point increase on the ACCESS writing scale score equals a 0.211 point increase of the PARCC reading scale score.

Using both correlation and multiple linear regression as statistical tests, the hypothesis is supported in that a student's English language proficiency in reading and writing impacts their reading achievement on the PARCC ELA standardized assessment to a significantly greater extent than their speaking and listening proficiencies.

The fourth overall hypothesis was: a student's English language proficiency in reading and writing will impact their mathematics achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies. Linear correlation and multiple linear regression were the statistical tests utilized to answer the second research question and test its underlying hypotheses.

For this hypothesis, the PARCC scale scores for math were the dependent variable, with the scale scores for each of the four ACCESS assessed language domains - speaking, listening, reading, and writing - as the independent variables. The primary assumptions for both correlation and multiple linear regression are similar: interval level variables, variables that are normally distributed, equal variances, variables that are independent of one another, and a linear relationship (Abbott, 2011). With respect to correlation, all assumptions were met with the exception of a randomly selected sample. Again, correlation is a robust test and can still provide meaningful results even with slight violations of its assumptions (Abbott, 2011). For multiple linear regression all of the assumptions were met. The variables are normally distributed and interval level. *Table16* shows the descriptive statistics for the study variables; the variables appear normal from the results.

Table 16

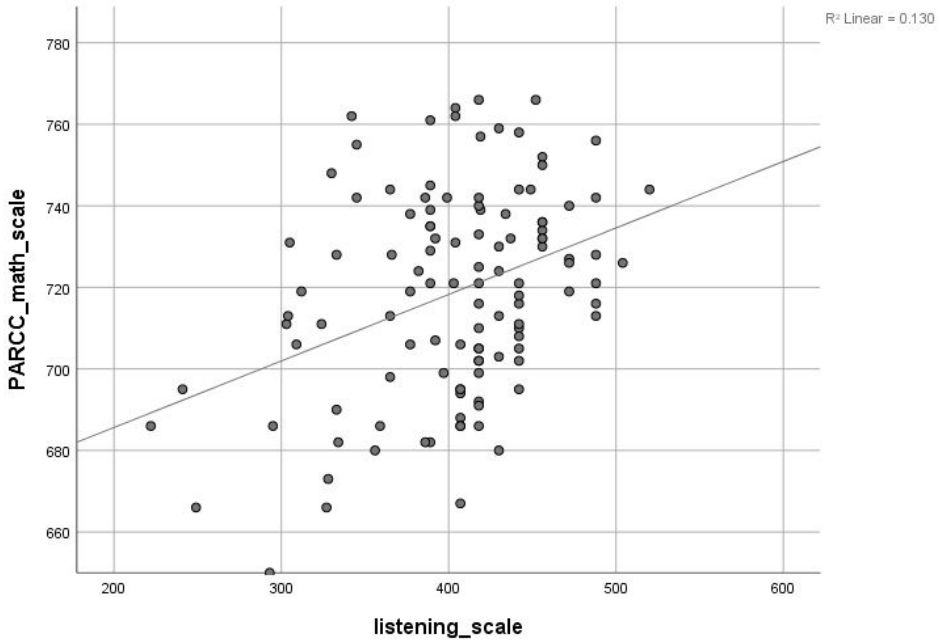
Descriptive Statistics

	N	Mean	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
PARCC math scale score	115	718.87	25.297	639.939	-0.234	0.226	-0.482	0.447
ACCESS listening scale score	119	402.660	55.384	3067.380	-0.750	0.222	0.806	0.440
ACCESS speaking scale score	119	325.670	48.578	2359.798	-1.312	0.222	2.921	0.440
ACCESS reading scale score	119	357.270	35.546	1263.520	-0.544	0.222	-0.440	0.440
ACCESS writing scale score	119	350.690	37.646	1417.233	-1.764	0.222	8.348	0.440

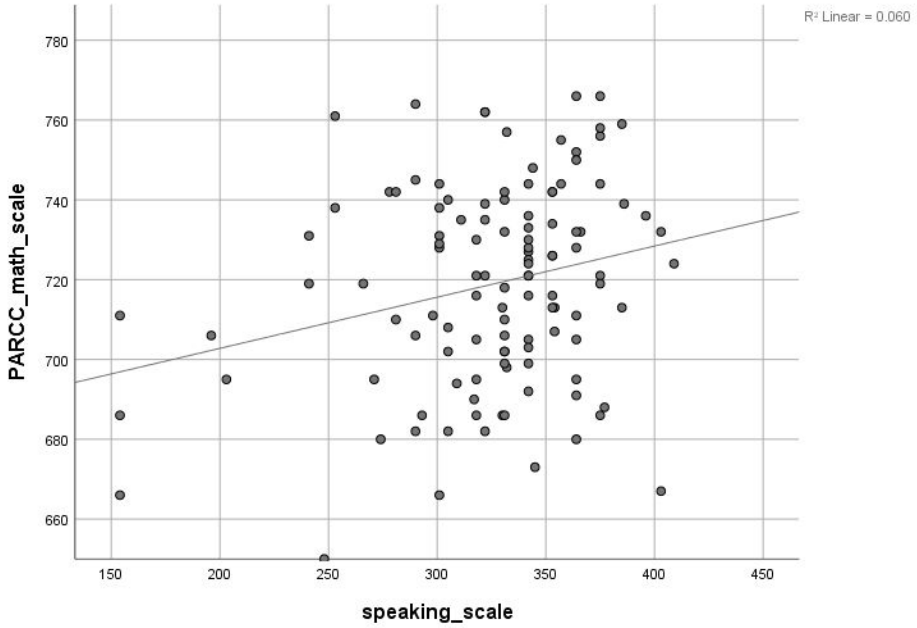
Table 16. Descriptive Statistics for Multiple Linear Regression of PARCC Math Scale Scores and ACCESS Language Domains.

Graph 4 depicts the histogram for PARCC math scores. *Graphs 6, 7, 8,* and *9* show the histograms for each of the four ACCESS assessed language domains - speaking, listening, reading, and writing. The variables appear normally distributed on each of the graphs. The variances can be considered equal as no graph is markedly skewed. *Graphs 14, 15, 16,* and *17* are scattergrams depicting PARCC math scale scores and each of the four ACCESS language

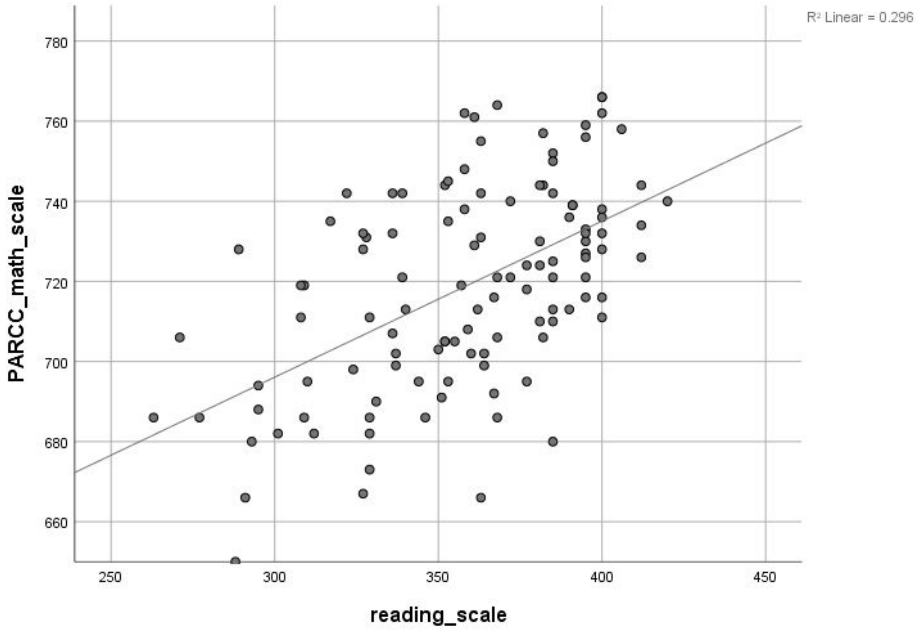
domains - listening, speaking, reading, and writing. Each graph illustrates, to varying degrees, a linear relationship between the variables. *Graphs 16* and *17* show the strongest correlation, illustrating the relationship between PARCC math scale scores and ACCESS reading and writing scale scores. The pattern of dots are fairly evenly distributed around the line of best fit, indicating a relatively strong and statistically significant relationship between these variables. The overall shape of the dots is an upward sloping straight line for both graphs. These visual patterns indicate a positive correlation. As a student's ACCESS reading and writing scale scores increase in value, their PARCC math scale score also increases. *Graph 14* and *15* show less significant, but still visible correlations between PARCC math scale scores and ACCESS listening and speaking scale scores. The overall cluster of dots is a bit broader, but still slopes upward reflecting a positive relationship between the variables.



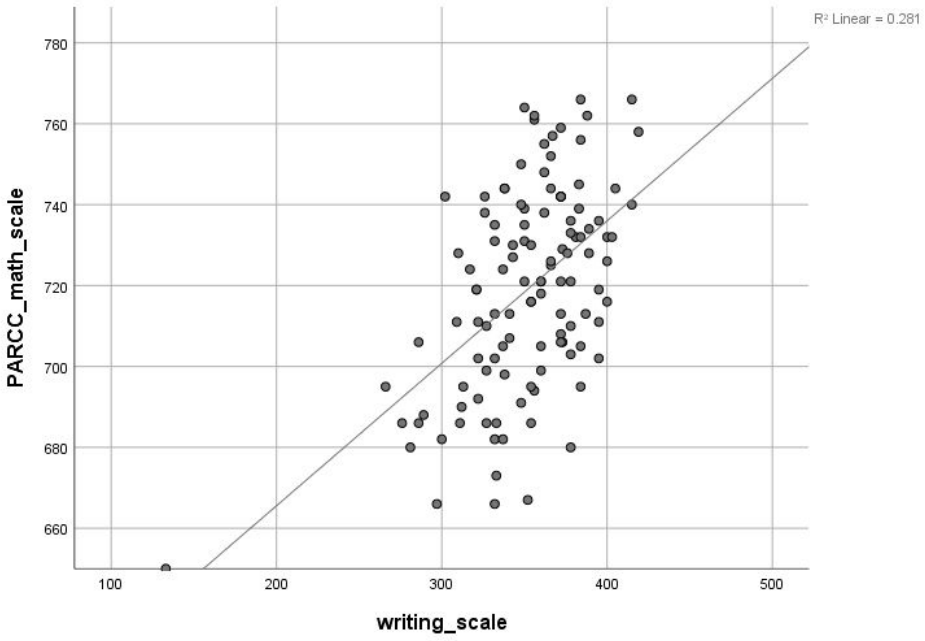
Graph 14. Scattergram of PARCC Math Scale Scores and ACCESS Listening Scale Scores.



Graph 15. Scattergram of PARCC Math Scale Scores and ACCESS Speaking Scale Scores.



Graph 16. Scattergram of PARCC Math Scale Scores and ACCESS Reading Scale Scores.



Graph 17. Scattergram of PARCC Math Scale Scores and ACCESS Writing Scale Scores.

Table 17 presents the data for the correlation of each of the four ACCESS assessed language domains and PARCC math scale scores. The following analysis employed three methods to determine the strength of the relationship between a student's proficiency in each of the four ACCESS assessed language domains - listening, speaking, reading, writing - and grade level content-area standardized assessments in math in response to the second research question and its hypothesis.

Table 17

Pearson Correlation

	PARCC math scale score	ACCESS listening scale score	ACCESS speaking scale score	ACCESS reading scale score	ACCESS writing scale score
PARCC math scale score	1.000	0.361	0.244	0.544	0.530
ACCESS listening scale score	0.361	1.000	0.660	0.718	0.605
ACCESS speaking scale score	0.244	0.660	1.000	0.535	0.540
ACCESS reading scale score	0.544	0.718	0.535	1.000	0.685
ACCESS writing scale score	0.530	0.605	0.540	0.685	1.000

Table 17. Pearson Correlation for PARCC Math Scale Scores and ACCESS Language Domains.

In consideration of ACCESS listening scale scores, the Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis (H0): p = 0, states that the relation between ACCESS listening scale scores and PARCC math scale scores is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 113 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would therefore be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.361$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.361 is statistically significant.

The second method for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.361 was transformed using the formula from Cohen (1988) to a *t* ratio of 4.113. The *T* table of values has a critical value of *t* 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.361) is significant because it resulted in a *t* score (4.113) that exceeded the critical *T* value (1.980).

The final method for assessing the strength of the relationship between the listening domain of language proficiency and mathematics achievement was to examine the effect size. According to Cohen (1988) a medium effect size has an r^2 value of 0.09 ($r = 0.30$) or greater, while a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.13 represents a medium effect size, indicating a relatively strong relationship between ACCESS listening scores and PARCC math scale scores.

With regard to ACCESS speaking scale scores, the Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis (H0): p = 0, states that the relation between ACCESS speaking scale scores and PARCC math scale score is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 113 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would therefore be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.244$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.244 is statistically significant.

The second method for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.244 was transformed using the formula from Cohen (1988) to a *t* ratio of 2.985. The *T* table of values has a critical value of *t* 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.244) is significant because it resulted in a *t* score (2.985) that exceeded the critical *T* value (1.980).

The final method for assessing the strengths of the relationship between the speaking domain of language proficiency and mathematics achievement was to examine the effect size. According to Cohen (1988) a small effect size has an r^2 value of 0.01 ($r = 0.10$) or greater; while a medium effect size has an r^2 value of 0.09 ($r = 0.30$) or greater. Using this measure our r^2 value of 0.60 represents a small effect size, indicating a small yet statistically significant relationship between ACCESS speaking scores and PARCC math scale scores.

With respect to ACCESS reading scale scores, the Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis (H0): p = 0, states that the relation between ACCESS reading scale scores and PARCC math scale scores is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 113 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would therefore be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.544$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.544 is statistically significant.

The second method of judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.544 was transformed using the formula from Cohen (1988) to a *t* ratio of 8.567. The *T* table of values has a critical value of *t* 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.544) is significant because it resulted in a *t* score (8.567) that exceeded the critical *T* value (1.980).

The final method for assessing the strengths of the relationship between the reading domain of language proficiency and mathematics achievement was to examine the effect size. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.296 represents a large effect size, indicating a strong and statistically significant relationship between ACCESS reading scores and PARCC math scale scores.

For ACCESS writing scale scores, the Hypothesis Test for Pearson's *r* was utilized first. The Null Hypothesis (H0): p = 0, states that the relation between ACCESS writing scale scores and PARCC math scale score is 0. The Alternative Hypothesis (HA): p does not equal 0, states that the correlation is not 0. The *Critical (Exclusion) Values for Pearson's Correlations Coefficient, r* table specifies that the two-tailed 0.05 critical value for *r* at 113 degrees of freedom is rdf(.05) = 0.195 (Appendix A). Values greater than 0.195 would therefore be considered statistically significant. The Null Hypothesis is rejected because the calculated value ($r = 0.530$) exceeds the critical value of exclusion (0.195). A significant correlation indicates that a calculated value as high as our value is unlikely to occur if there is no correlation between the data sets. Thus, the calculated *r* of 0.530 is statistically significant.

The second method for judging the strength of the relationship between the two variables was by transforming the *r* value into a *t* ratio and comparing it against the actual values of *T* outlined in the *Exclusion Values for the T Distribution* table (Appendix B). The *r* value of 0.530 was transformed using the formula from Cohen (1988) to a *t* ratio of 8.213. The *T* table of values has a critical value of *t* 0.05, 120 = 1.980. Therefore our calculated value of *r* (0.530) is significant because it resulted in a *t* score (8.213) that exceeded the critical *T* value (1.980).

The final method for assessing the strengths of the relationship between the written domain of language proficiency and mathematics achievement was to examine the effect size. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.281 represents a large effect size, indicating a strong relationship between ACCESS writing scores and PARCC math scale scores.

Table 18 displays the summary data for the linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.336, equates to approximately 34% of the variance in PARCC math scores explained by the four language domains of the ACCESS assessment - listening, speaking, reading, and writing.

Table 18

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.599*	0.359	0.336	20.616	0.359	15.412	4	110	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale

Table 18. Multiple Linear Regression Model Summary.

Table 19 displays the ANOVA summary table for the multiple linear regression model. The *F* test shows a value of 15.412 with a significance of 0.000, less than the 0.05 threshold to be considered statistically significant.

Table 19

*ANOVA**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26200.965	4	6550.241	15.412	.000**
	Residual	46752.079	110	425.019		
	Total	72953.043	114			

* Dependent Variable: PARCC_math_scale

** Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale

Table 19. ANOVA Summary.

Table 20 depicts the coefficient outputs for the regression model. When examining the coefficient data of the multiple linear regression model, the impact of each ACCESS language proficiency on the PARCC math scale scores can be seen. The ACCESS reading scale score, with a *p*-value of 0.001, is statistically significant. The ACCESS writing scale score is also significant with a *p*-value of 0.002. Both the ACCESS listening and speaking scale scores are not significant with *p*-values of 0.556 and 0.242 respectively.

Table 20

Coefficients

Model		Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B	
							Lower Bound	Upper Bound
1	(constant)	562.928	20.784		27.085	0.000	521.739	604.118
	ACCESS listening scale score	-0.033	0.057	-0.074	-0.591	0.556	-0.145	0.079
	ACCESS speaking scale score	-0.065	0.055	-0.123	-1.176	0.242	-0.173	0.044
	ACCESS reading scale score	0.302	0.088	0.422	3.426	0.001	0.127	0.477
	ACCESS writing scale score	0.234	0.073	0.352	3.203	0.002	0.089	0.379

* Dependent Variables: PARCC_math_scale

Table 20. Coefficient Outputs.

Accordingly, utilizing the unstandardized *B* coefficients the model predicts a one point increase on the ACCESS reading scale score equals a 0.302 point increase of the PARCC math scale score. A one point increase on the ACCESS writing scale score equals a 0.234 point increase of the PARCC math scale score.

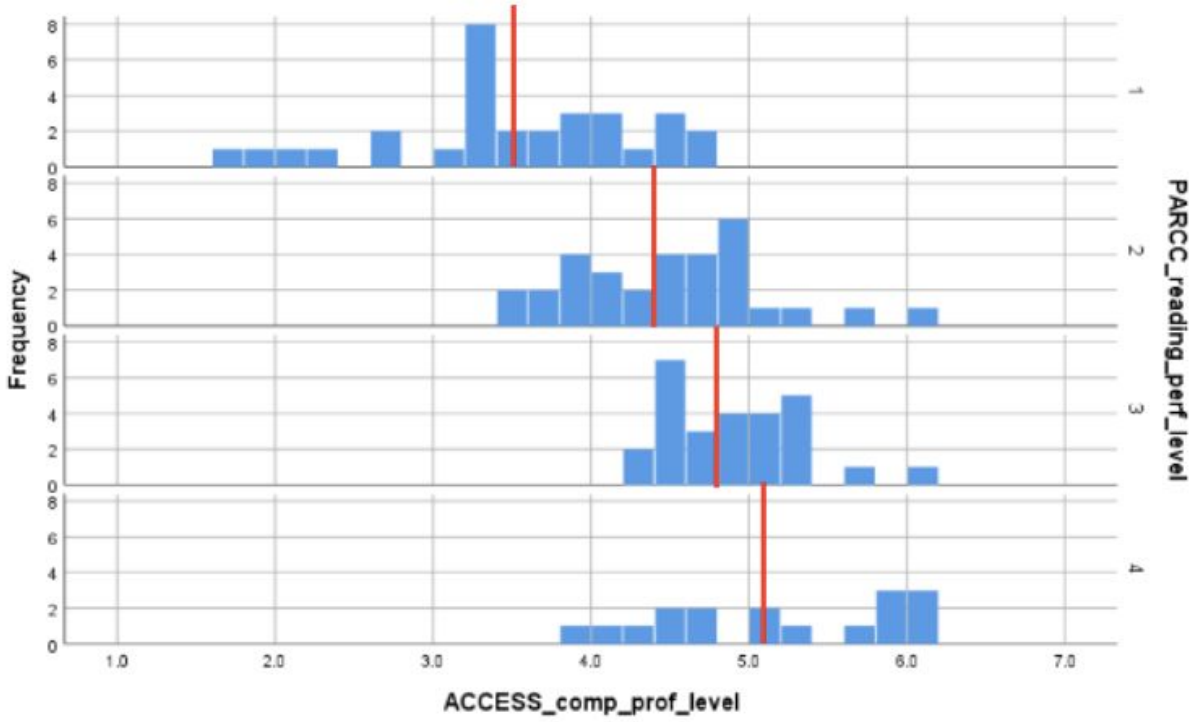
Using both correlation and multiple linear regression as statistical tests, the hypothesis is supported in that a student's English language proficiency in reading and writing impacts their mathematics achievement on standardized assessment to a significantly greater extent than their speaking and listening proficiencies.

Research Question Three

The third research question guiding this study was: what is the threshold English language proficiency for ELs to meet/exceed standards on federally mandated standardized assessments in ELA and mathematics? The hypothesis pertaining to this question was: according to Illinois school code ELs are defined as proficient at a composite ACCESS score of 4.8 or beyond, and should thus be able to demonstrate proficiency on grade level content-area assessments.

To answer this research question and test its hypothesis, mean and median ACCESS composite scores were compared within the ranges of PARCC academic proficiency levels. This treatment of the data allowed for determining the overall composite ACCESS score that serves as the threshold for achieving proficiency on the PARCC ELA and mathematics assessment for our particular student population. Histograms and cross tabulation tables are used to visually illustrate the data.

Graph 18 is a histogram detailing the frequency of composite ACCESS proficiency levels by each of the PARCC proficiency levels for ELA.



Graph 18. Histogram of Composite ACCESS Proficiency Levels by PARCC ELA Proficiency Level

No students in the study scored at a performance level 5 (exceeded standards). The red line indicates the mean English language proficiency level for each of the PARCC performance

levels. The increase in mean language proficiency as students perform at higher levels on the PARCC ELA assessment is clearly illustrated. Hence, the relationship between reading achievement and English language proficiency is further reinforced by this graph and accompanying analysis of the data.

Beginning with the PARCC ELA data and the ACCESS scores needed to meet/exceed grade level reading standards, the mean ACCESS composite proficiency level within the sample student population to achieve a performance level 4 (met standards) on the PARCC ELA assessment was a 5.1. The median score of the same data set was 5.1. The range of ACCESS composite proficiency levels within this tier of meeting standards was 3.8 to 6.0. By comparison the mean ACCESS composite proficiency level for students who achieved at a performance level 3 (approached standards) was a 4.8, with 4.8 as the median. The range of ACCESS composite proficiency levels for this performance level was 4.2 to 6.0. The mean ACCESS composite proficiency level for students who achieved at a performance level 2 (partially met standards) was 4.4, with a median score of 4.4. Students performing at a level 2 on the PARCC ELA assessment represented a wider range of English language proficiency than the previous two performance levels, ranging from 3.4 to a 6. The mean ACCESS composite proficiency level for students who achieved at a performance level 1 (partially met standards) was 3.5, with a median score of 3.4. Students performing at a level 1 also represented a wider range of English language proficiency than higher performing levels, ranging from 1.7 to 4.7.

Table 21 shows the breakdown of achievement on the PARCC ELA assessment by English and PARCC proficiency levels. The horizontal red line delineates the cut score (750) for students to be deemed "meeting/exceeding" grade level expectations for reading on the PARCC

assessment. The vertical red line indicates the approximate point within the data that the state reclassifies ELs based upon language proficiency (4.8).

Table 21

Reading achievement of EL students and English language proficiency

	English Language Proficiency			
Reading Proficiency	Low (1.0-2.9)	Medium (3.0-4.9)	High (5.0+)	Total percentage
	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)
	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)
Level 1: did not yet meet expectations (650-699)	19.4 (6)	80.6 (25)	0 (0)	100 (31)
	100 (6)	33.3 (25)	0 (0)	29.2 (31)
Level 2: partially met expectations (700-724)	0 (0)	87.1 (27)	12.9 (4)	100 (31)
	0 (0)	36 (27)	16 (4)	29.2 (31)
Level 3: approached expectations (725-749)	0 (0)	59.3 (16)	40.7 (11)	100 (27)
	0 (0)	21.3 (16)	44 (11)	25.5 (27)
Level 4: met expectations (750-809)	0 (0)	41.2 (7)	58.5 (10)	100 (17)
	0 (0)	9.3 (7)	40 (10)	16 (17)
Level 5: exceeded expectations (810-850)	0 (0)	0 (0)	0 (0)	0 (0)
	0 (0)	0 (0)	0 (0)	0 (0)
Total (N= 106)	5.7 (6)	70.8 (75)	23.6 (25)	100 (106)
	100 (6)	100 (75)	100 (25)	100 (106)

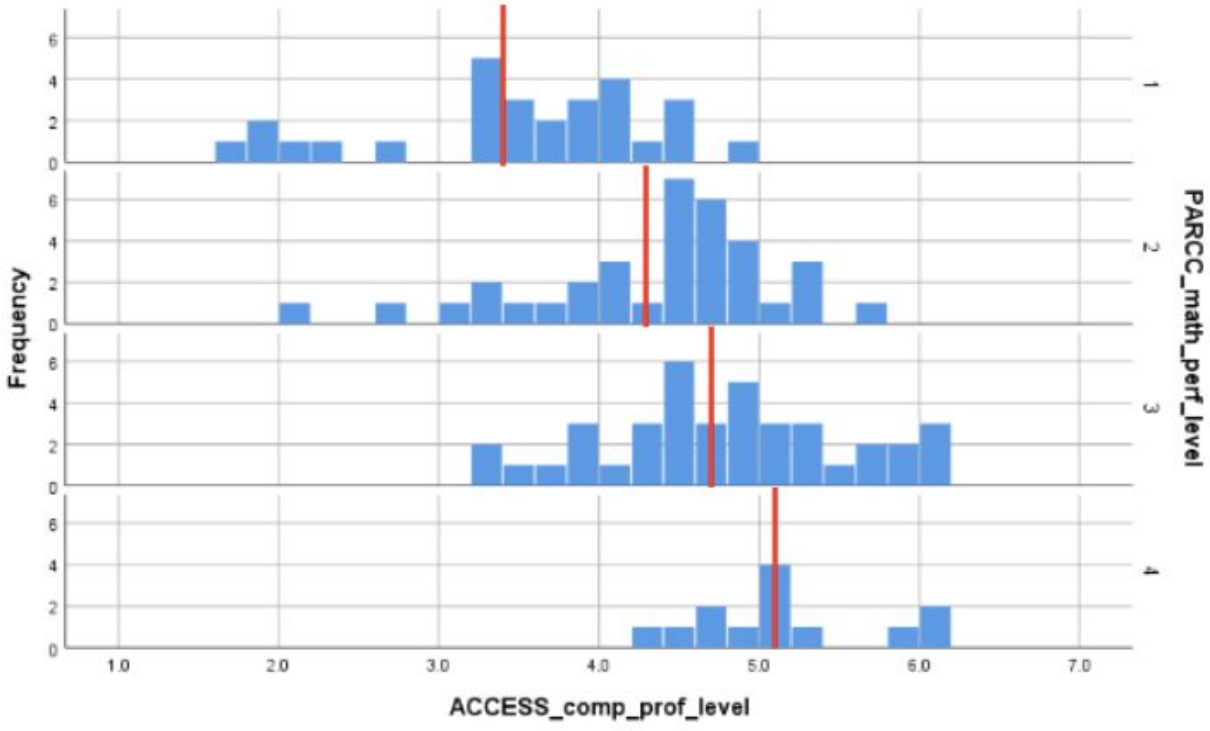
Table 21. Contingency Table for English Proficiency Levels and Grade Level Reading Proficiency Levels.

16% of students in the study performed at a level 4 (met expectations) for the PARCC ELA assessment, none at a level 5 (exceeded expectations). This was substantially above the average performance of ELs within the state, with only 11% scoring at a level 4 and 0% at a level 5 for the same 2018 PARCC ELA assessment (Illinois Report Card, 2019). Also illustrated in the table is that 100% of students with low English proficiency scored at a level 1 (did not yet meet expectations). While 58.5% of students with high English proficiency scored at a level that was equivalent to meeting grade level standards for reading, not a single student with low English proficiency met that benchmark. Furthermore no students with high English proficiency levels

scored at a level 1. *Table 21* again reinforces the significant relationship between reading achievement on standardized assessments and English language proficiency for ELs within this student population.

Based on the student population sampled the threshold language proficiency for ELs to meet standards on the PARCC ELA assessment was a composite ACCESS proficiency level of 5.1. This score equates to the "bridging" level of language proficiency. This score is beyond the 4.8 threshold the state uses to reclassify ELs and demonstrates a misalignment between research, policy, and assessment practices.

Graph 19 is a histogram detailing the frequency of composite ACCESS proficiency levels by each of the PARCC proficiency levels for math.



Graph 19. Histogram of Composite ACCESS Proficiency Levels by PARCC Math Proficiency Level

No students in the study scored at a performance level 5 (exceeded standards). The red line indicates the mean English language proficiency level for each of the PARCC performance levels. The increase in mean language proficiency as students perform at higher levels on the PARCC math assessment is clearly illustrated. Hence, the relationship between mathematics achievement and English language proficiency is further reinforced by this graph and accompanying analysis of the data.

Examining the PARCC mathematics data and the ACCESS scores needed to meet/exceed grade level math standards, the mean ACCESS composite proficiency level within the sample student population to achieve a performance level 4 (met standards) on the PARCC math assessment was a 5.1. The median score of the same data set was 5.1. The range of ACCESS composite proficiency levels within this tier of meeting standards was 4.2 to 6.0. By comparison the mean ACCESS composite proficiency level for students who achieved at a performance level 3 (approached standards) was a 4.7, with 4.7 as the median. The range of ACCESS composite proficiency levels for this performance level was 3.2 to 6.0. The mean ACCESS composite proficiency level for students who achieved at a performance level 2 (partially met standards) was 4.3, with a median score of 4.5. Students performing at a level 2 on the PARCC math assessment represented a wider range of English language proficiency than the previous two performance levels, ranging from 2.1 to 5.7. The mean ACCESS composite proficiency level for students who achieved at a performance level 1 (partially met standards) was 3.4, with a median score of 3.6. Students performing at a level 1 also represented a wider range of English language proficiency than higher performing levels, ranging from 1.7 to 4.8.

Table 22 shows the breakdown of achievement on the PARCC math assessment by English and PARCC proficiency levels. The horizontal red line delineates the cut score (750) for students to be deemed "meeting/exceeding" grade level expectations for reading on the PARCC assessment. The vertical red line indicates the approximate point within the data that the state reclassifies ELs based upon their language proficiency (4.8).

Table 22

Mathematics achievement of EL students and English language proficiency

	English Language Proficiency			
Mathematics Proficiency	Low (1.0-2.9)	Medium (3.0-4.9)	High (5.0+)	Total percentage
	row percentage (N)	row percentage (N)	row percentage (N)	row percentage (N)
	column percentage (N)	column percentage (N)	column percentage (N)	column percentage (N)
Level 1: did not yet meet expectations (650-699)	21.4 (6)	78.6 (22)	0 (0)	100 (28)
	75 (6)	27.5 (22)	0 (0)	24.3 (28)
Level 2: partially met expectations (700-724)	5.7 (2)	80 (28)	14.3 (5)	100 (35)
	25 (2)	35 (28)	18.5 (5)	30.4 (35)
Level 3: approached expectations (725-749)	0 (0)	64.1 (25)	35.9 (14)	100 (39)
	0 (0)	31.3 (25)	51.9 (14)	34.0 (39)
Level 4: met expectations (750-809)	0 (0)	38.5 (5)	61.5 (8)	100 (13)
	0 (0)	6.3 (5)	29.6 (8)	11.3 (13)
Level 5: exceeded expectations (810-850)	0 (0)	0 (0)	0 (0)	0 (0)
	0 (0)	0 (0)	0 (0)	0 (0)
Total (N= 115)	7.0 (8)	69.5 (80)	23.5 (27)	100 (115)
	100 (8)	100 (80)	100 (27)	100 (115)

Table 22. Contingency Table for English Proficiency Levels and Grade Level Math Proficiency Levels.

11.3% of students in the study performed at a level 4 (met expectations) for the PARCC math assessment, none at a level 5 (exceeded expectations). This was about the same as the average performance of ELs within the state, with 11% scoring at a level 4 and 1% at a level 5 for the same 2018 PARCC math assessment (Illinois Report Card, 2019). Also illustrated in the table is that 100% of students with low English proficiency scored at a level 1 (did not yet meet

expectations) or level 2 (partially met expectations). While 61.5% of students with high English proficiency scored at a level that was equivalent to meeting grade level standards for math. Furthermore no students with high English proficiency levels scored at a level 1. *Table 23* again reinforces the significant relationship between mathematics achievement and English language proficiency for ELs within this student population.

Based on the student population sampled the threshold language proficiency for ELs to meet standards on the PARCC math assessment was a composite ACCESS proficiency level of 5.1. This score equates to the "bridging" level of language proficiency. This score is beyond the 4.8 threshold the state uses to reclassify ELs and demonstrates a misalignment between research, policy, and assessment practices.

Based upon analysis of the mean and median ACCESS composite scores in comparison to the ranges of PARCC academic proficiency levels, the threshold language proficiency for ELs to meet standards on the PARCC assessment for both ELA and Mathematics is at a composite ACCESS proficiency level of 5.1. This score equates to the "bridging" level of language proficiency and is beyond the 4.8 threshold the state uses to reclassify ELs. Hence the point the state defines as ELs no longer needing, and providing financial support for, targeted language instruction, the data reveals as inadequate relative to achieving at the state's own cut scores for demonstrating grade level proficiency in reading and math. This data further reveals the misalignment between research, policy, and assessment practices for ELs.

Research Question Four

The fourth research question guiding this study was: what is the impact of other student demographic factors - race/ethnicity, socioeconomic status - in relation to English language

proficiency for the achievement of ELs on federally mandated standardized assessments in ELA and mathematics? Seven linear regression models were used to answer this final research question and test its two accompanying hypotheses. The use of multiple linear regressions allows for the comparison of model summary changes with the introduction of additional predictor variables (Abbott, 2011). Overall, PARCC scale scores for ELA and mathematics were the dependent (outcome) variable and English language proficiency level, language domains - speaking, listening, reading, and writing, socioeconomic status, and race/ethnicity the independent variables. The first hypothesis for this research question was: controlling for other variables, ELs identifying with a race/ethnicity of White will achieve higher levels of reading and mathematics proficiency. The second hypothesis for this question was: controlling for other variables, ELs with full pay lunch status will achieve higher levels of reading and mathematics proficiency.

As was discussed earlier in this chapter, the primary assumptions for correlation as a statistical procedure are: a randomly chosen sample, variables that are independent of one another, variables that are normally distributed, equal variances, and a linear relationship (Abbott, 2011). In this study, all assumptions with the exception of a randomly chosen sample were met. The data sets were not linked and thus independent of one another. Both data sets represented scaled scores and were interval. *Tables 23* and *24* show the descriptive statistics for the logistic regression analysis; the variables appear to be normal from the results.

Table 23				
Descriptive Statistics for Variables in the Analysis				
	Mean	Std. Deviation	Variance	Range
PARCC ELA Scale Score	716.540	31.497	992.080	650 to 780
PARCC Math Scale Score	718.870	25.297	639.939	650 to 766
Low English Language Proficiency	0.067	0.251	0.063	0 to 1
Medium English Language Proficiency	0.706	0.458	0.209	0 to 1
High English Language Proficiency	0.227	0.227	0.177	0 to 1
ACCESS Listening Scale Score	402.660	55.384	3067.380	222 to 520
ACCESS Speaking Scale Score	325.670	48.578	2359.798	154 to 429
ACCESS Reading Scale Score	357.270	35.546	1263.520	263 to 420
ACCESS Writing Scale Score	350.690	37.646	1417.233	133 to 419
Free Lunch Status	0.656	0.477	0.228	0 to 1
Reduced Lunch Status	0.101	0.302	0.091	0 to 1
Full Pay Lunch Status	0.244	0.431	0.186	0 to 1
Asian	0.429	0.497	0.247	0 to 1
Black	0.0420	0.201	0.041	0 to 1
Hispanic	0.152	0.360	0.129	0 to 1
Other	0.017	0.129	0.017	0 to 1
White	0.361	0.482	0.233	0 to 1
Sample Size n = 119				

Table 23. Descriptive Statistics for Variables in Regression Analysis.

Table 24				
Descriptive Statistics for Variables in the Analysis				
	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
PARCC ELA Scale Score	-0.199	0.235	-0.482	0.465
PARCC Math Scale Score	-0.234	0.226	-0.482	0.447
Low English Language Proficiency	3.501	0.222	10.430	0.440
Medium English Language Proficiency	-0.915	0.222	-1.182	0.440
High English Language Proficiency	1.321	0.222	-0.260	0.440
ACCESS Listening Scale Score	-0.750	0.222	0.806	0.440
ACCESS Speaking Scale Score	-1.312	0.222	2.921	0.440
ACCESS Reading Scale Score	-0.544	0.222	-0.440	0.440
ACCESS Writing Scale Score	-1.764	0.222	8.348	0.440
Free Lunch Status	-0.663	0.222	-1.588	0.440
Reduced Lunch Status	2.685	0.222	5.299	0.440
Full Pay Lunch Status	1.209	0.222	-0.547	0.440
Asian	0.292	0.222	-1.948	0.440
Black	4.624	0.222	19.712	0.440
Hispanic	1.972	0.222	1.919	0.440
Other	7.614	0.222	56.931	0.440
White	0.585	0.222	-1.687	0.440
Sample Size n = 119				

Table 24. Descriptive Statistics for Variables in Regression Analysis.

Seven logistic regression models were estimated using the logarithm of the PARCC scale scores for ELA and math as the dependent variable and English language proficiency level, ACCESS assessed language domains - speaking, listening, reading, and writing, socioeconomic status, and race/ethnicity as independent variables. These models tested the statistical significance of the effects of the independent variables on PARCC ELA and math performance. Reading achievement was modeled first. *Table 25* shows the results for the estimation of the parameters for the seven logistic regression models in relation to PARCC ELA scores.

Table 25							
Linear Regression Models for PARCC Reading Scale Scores							
Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient
	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error
Medium English Language Proficiency	44.133**				-0.238	-1.429	-5.618
	10.698				13.979	13.748	13.798
High English Language Proficiency	79.053**				10.929	11.365	5.171
	11.463				17.975	17.668	17.948
ACCESS Listening Scale Score		-0.062			-0.073	-0.053	-0.030
		0.064			0.065	0.064	0.065
ACCESS Speaking Scale Score		0.022			0.023	0.014	0.027
		0.063			0.066	0.065	0.066
ACCESS Reading Scale Score		0.524**			0.464**	0.434**	0.413**
		0.101			0.110	0.109	0.109
ACCESS Writing Scale Score		0.211**			0.196**	0.203**	0.200**
		0.081			0.088	0.086	0.087
Reduced Lunch Status			13.629			17.325**	16.807**
			10.601			7.672	7.831
Full Pay Lunch Status			9.705			5.588	8.176
			7.202			5.160	5.293
Asian				-14.286			-21.469**
				14.304			10.469
Hispanic				-25.047			-22.724**
				15.419			11.378
White				-34.847**			-26.039**
				14.419			10.813
Other				-22.400			-31.810*
				25.358			19.122
Constant	666.667**	472.304**	712.871**	739.400**	500.676**	501.432**	522.743***
	10.294	24.556	3.748	13.554	34.115	33.586	34.898

p-value* < or = 0.10; p-value** < or = 0.05

Table 25. Summary of Linear Regression Models for PARCC ELA Scale Scores.

Model one included English language proficiency variables only. The PARCC scale scores for ELA were the dependent variable, with English language proficiency levels the independent variables. ELs were divided into three levels measuring different degrees of English language proficiency: low, medium, and high. This variable was operationalized by creating dummy variables for each distinct level of language proficiency. Students with low English language proficiency served as the reference category for the first model.

Table 26

Model One Summary - Reading

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	0.609*	0.371	0.359	25.215	0.371	30.420	2	103	0.000

* Predictors: (Constant), high, medium

Table 26. Model One Summary - Reading.

Table 26 displays the summary data for the first linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.359, equates to approximately 36% of the variance in PARCC reading scores explained by a student's level of English language proficiency. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.359 represents a large effect size, indicating a strong relationship between English language proficiency and PARCC ELA performance.

Table 25 shows the results for the estimation of the parameters. When examining the unstandardized *B* coefficient data of the multiple linear regression model, the impact of a student's English language proficiency on the PARCC ELA scale scores can be seen. Both language proficiency variables are statistically significant. Alone in the model, medium and high English language proficiency are both significant with a *p*-value of 0.000. Accordingly, the regression analysis predicts that increasing a student's English language proficiency to the medium range results in an PARCC ELA score increase of 44 points, while moving a student to high English language proficiency results in an increase of 79 points. Based on PARCC's scaled scoring system, this is equivalent to a one performance level increase for each added range of English language proficiency.

Model two included the four ACCESS assessed language domains of speaking, listening, reading, and writing only. The PARCC scale scores for ELA were the dependent variable, with the scaled scores for each of the four language domains as the independent variables.

Table 27

Model Two Summary - Reading

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
2	0.710*	0.504	0.485	22.611	0.504	25.685	4	101	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale

Table 27. Model Two Summary - Reading.

Table 27 displays the summary data for the second linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.485, equates to approximately 49% of the variance in PARCC reading scores explained by a student's level of English proficiency in the domains of speaking, listening, reading, and writing. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.485 represents a large effect size, indicating a strong relationship between English language proficiency and PARCC ELA performance.

Table 25 shows the results for the estimation of the parameters. When examining the unstandardized *B* coefficient data of the multiple linear regression model, the impact of a student's English language proficiency on the PARCC ELA scale scores can be seen. Both the reading and writing variables are statistically significant. Reading had a *p*-value of 0.000, and writing a *p*-value of 0.011. The speaking and listening variables were not statistically significant. These findings align with those of the second research question. In addition, the regression model predicts that a one point increase of a student's ACCESS reading subscore will result in a half-point increase on the PARCC ELA assessment. It also predicts that a one point

increase of a student's ACCESS writing subscore will result in a one-fifth point increase on the PARCC ELA assessment.

Model three introduced socioeconomic status variables. The PARCC scale scores for ELA were the dependent variable, with a student's lunch status serving as a proxy for socioeconomic status as the independent variables. ELs were divided into three levels measuring different degrees of socioeconomic status: free lunch, reduced lunch, and full pay lunch. This variable was operationalized by creating dummy variables for each distinct level. Students with free lunch served as the reference category for the third model.

Table 28

Model Three Summary - Reading

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
3	.166*	0.028	0.009	31.359	0.028	1.465	2	103	0.236

* Predictors: (Constant), pay, reduced

Table 28. Model Three Summary - Reading.

Table 28 displays the summary data for the third linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.009, equates to less than one tenth of one percent of the variance in PARCC ELA scores explained by a student's socioeconomic status. According to Cohen (1988) a small effect size has an r^2 value of 0.01 ($r = 0.10$) or greater. Using this measure our r^2 value of 0.009 is not statistically significant and reflects no meaningful relationship between a student's socioeconomic status and PARCC ELA performance.

Table 25 shows the results for the estimation of the parameters. Neither the reduced or full pay variables were statistically significant within the model. According to model three there is not a statistically significant relationship between socioeconomic status and reading

proficiency as measured by PARCC for the student population studied. This finding rejects the hypothesis that full pay students will perform better on this standardized assessment than their lower socioeconomic peers.

Model four included race/ethnicity variables only. The PARCC scale scores for ELA were the dependent variable, with race/ethnicity as the independent variables. ELs self-identified into one of five racial/ethnic categories: Asian, Black, Hispanic, White, and Other. This variable was operationalized by creating dummy variables for each distinct category. Students identifying as Black served as the reference category for the fourth model.

Table 29

Model Four Summary - Reading

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
4	.331*	0.109	0.074	30.308	0.109	3.100	4	101	0.019

* Predictors: (Constant), asian, hispanic, white, other

Table 29. Model Four Summary - Reading.

Table 29 displays the summary data for the fourth linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.074, equates to approximately 7% of the variance in PARCC ELA scores explained by a student's race/ethnicity. According to Cohen (1988) a small effect size has an r^2 value of 0.01 ($r = 0.10$) or greater. Using this measure our r^2 value of 0.074 is not statistically significant and reflects little relationship between a student's race/ethnicity and PARCC ELA performance.

Table 25 shows the results for the estimation of the parameters. When examining the unstandardized *B* coefficient data of the multiple linear regression model, the impact of a student's race/ethnicity on the PARCC ELA scale scores can be seen. Identification as White was the only racial/ethnic variable of statistical significance. The White variable had a *p*-value

of 0.017. Accordingly, the regression analysis predicts that White students will score approximately 35 points lower on the PARCC ELA assessment than students identifying as Black. This finding rejects the hypothesis that White students would perform better on this standardized assessment than other racial/ethnic groups.

Model five included English language proficiency variables and the four ACCESS assessed language domains of speaking, listening, reading, and writing only. There exists a degree of overlap within this model, as a weighted combination of the four language domains - 35% Reading + 35% Writing + 15% Listening + 15% Speaking - was used to determine the composite scale scores that define the English language proficiency classifications. The PARCC scale scores for ELA were the dependent variable, with English language proficiency levels and the scaled scores for each of the four language domains as the independent variables.

Table 30

Model Five Summary - Reading

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
5	.720*	0.519	0.489	22.505	0.519	17.779	6	99	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale, medium, high

Table 30. Model Five Summary - Reading.

Table 30 displays the summary data for the fifth linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.489, equates to approximately 49% of the variance in PARCC ELA scores explained by a student's level of English language proficiency in reading, writing, speaking, and listening. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.489 represents a large effect size, indicating a significant relationship between these English language proficiency variables and PARCC ELA performance.

Table 25 shows the results for the estimation of the parameters. With the introduction of these language proficiency variables together, a student's level of English proficiency as medium or high is no longer statistically significant. The reading and writing language variables are still statistically significant with *p*-values of 0.000 and 0.027 respectively. Accordingly, the regression analysis prediction remains relatively constant in that a one point increase on a student's ACCESS reading subscore will result in a half-point increase on the PARCC ELA assessment. It also predicts that a one point increase on a student's ACCESS writing subscore will result in a one-fifth point increase on the PARCC ELA assessment. This model again underscores the link between a student's English language proficiency in reading and writing and increased performance on the PARCC ELA assessment.

Model six introduced socioeconomic variables. The PARCC scale scores for ELA were the dependent variable, with English language proficiency levels, scaled scores for each of the four ACCESS assessed language domains, and socioeconomic status as the independent variables.

Table 31

Model Six Summary - Reading

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
6	.738*	0.545	0.507	22.105	0.545	14.524	8	97	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale, medium, high, reduced, pay

Table 31. Model Six Summary - Reading.

Table 31 displays the summary data for the sixth linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.507, equates to approximately 51% of the variance in PARCC ELA scores explained by the combination of these variables. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using

this measure our r^2 value of 0.507 represents a large effect size, indicating a significant relationship between the independent variables and PARCC ELA performance.

Table 25 shows the results for the estimation of the parameters. With the introduction of the socioeconomic variables, a student's level of English proficiency as medium or high continues to no longer be statistically significant. The reading and writing language variables remain statistically significant with *p*-values of 0.000 and 0.020 respectively. The predictive impact of each variable also remains relatively constant. In this model, the reduced lunch variable becomes statistically significant with a *p*-value of 0.026. The full pay lunch variable is not. The model predicts that students moving from free to reduced lunch status will increase their PARCC ELA score by approximately 17 points. As a result, socioeconomic status is an impacting factor within the model, though it appears to be limited with a ceiling at full pay lunch status. Model six further rejects the socioeconomic hypothesis that full pay students will perform at higher levels, but does indicate a relationship between poverty and lower levels of performance on the PARCC ELA assessment.

Model seven introduced race/ethnicity variables. The PARCC scale scores for ELA were the dependent variable, with English language proficiency levels, scaled scores for each of the four ACCESS assessed language domains, socioeconomic status, and race/ethnicity as the independent variables.

Table 32

Model Seven Summary - Reading

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
7	.757*	0.573	0.518	21.870	0.573	10.399	12	93	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale, medium, high, reduced, pay. asian, hispanic, white, other

Table 32. Model Seven Summary - Reading.

Table 32 displays the summary data for the seventh linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.518, equates to approximately 52% of the variance in PARCC ELA scores explained by the combination of these variables. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.518 represents a large effect size, indicating a significant relationship between the independent variables and PARCC ELA performance.

Table 25 shows the results for the estimation of the parameters. With the introduction of the racial/ethnic variables, a student's level of English proficiency as medium or high remains unchanged and not statistically significant. With the introduction of multiple variables into later models, English proficiency levels shifted from being significant to a high degree by itself, to less impacting with the other compounding variables. The reading and writing language variables remain statistically significant with *p*-values of 0.000 and 0.024 respectively. This significance remains constant throughout each of the models. The predictive impact of each variable also remains relatively constant, 0.413 for reading and 0.200 for writing. The reduced lunch variable continues to be statistically significant with a *p*-value of 0.034, with its predictive impact approximately the same. The full pay lunch variable remains not statistically significant. In this model, each racial/ethnic variable is statistically significant. The Asian variable has a *p*-value of 0.043; Hispanic of 0.049, White of 0.018, and Other of 0.010. The predictive impact of each race/ethnicity variable was negative: approximately 21 points for Asian, 23 points for Hispanic, 26 points for White, and 32 points for Other in comparison to the Black students that served as the reference group.

Accordingly, socioeconomic status is an impacting factor within the model, though it appears to be limited with a ceiling at full pay status. Model seven again rejects the socioeconomic hypothesis that full pay students will perform at higher levels, but continues to indicate a relationship between poverty and decreased performance on the PARCC ELA assessment. Model seven also rejects the racial/ethnic hypothesis that students identified as White will perform at higher levels on this standardized reading assessment. The inclusion of race/ethnicity variables within the model makes each a statistically significant factor, and refutes the hypothesis as each group is predicted to perform at a lower level than the Black reference group.

The same seven logistic regression models were used to analyze mathematics performance. *Table 33* shows the results for the estimation of the parameters for the seven logistic regression models with respect to PARCC math scores.

Table 33							
Linear Regression Models for PARCC Mathematics Scale Scores							
Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient	Unstandardized B Coefficient
	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error	Coefficients Std. Error
Medium English Language Proficiency	30.488**				18.052	16.607	12.293
	7.990				11.651	11.411	11.248
High English Language Proficiency	53.926**				30.927**	32.116**	25.216*
	8.674				15.167	14.842	14.783
ACCESS Listening Scale Score		-0.033			-0.052	-0.039	-0.010
		0.057			0.056	0.056	0.055
ACCESS Speaking Scale Score		-0.065			-0.096	-0.105*	-0.096*
		0.055			0.058	0.057	0.057
ACCESS Reading Scale Score		0.302**			0.222**	0.215**	0.189**
		0.088			0.094	0.093	0.091
ACCESS Writing Scale Score		0.234**			0.178**	0.172**	0.180**
		0.073			0.077	0.076	0.076
Reduced Lunch Status			12.874			15.847**	15.462**
			8.484			6.868	6.902
Full Pay Lunch Status			-1.133			-3.925	-1.982
			5.570			4.445	4.441
Asian				-20.612*			-23.307**
				11.155			9.233
Hispanic				-30.967**			-28.243**
				12.000			10.004
White				-37.776**			-26.755**
				11.230			9.471
Other				-39.300**			-39.732**
				19.861			16.851
Constant	685.000**	562.928**	718.026**	746.800**	608.908**	612.301**	632.382**
	7.618	20.784	2.876	10.616	29.065	28.450	29.510

p-value* < or = 0.10; p-value** < or = 0.05

Table 33. Summary of Linear Regression Models for PARCC Math Scale Scores.

Model one included English language proficiency variables only. The PARCC scale scores for mathematics were the dependent variable, with English language proficiency levels the independent variables. ELs were again divided into three levels measuring different degrees of English language proficiency: low, medium, and high. This variable was operationalized by creating dummy variables for each distinct level of language proficiency. Students with low English language proficiency served as the reference category for the first model.

Table 34

Model One Summary - Mathematics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.536*	0.287	0.274	21.547	0.287	22.565	2	112	0.000

* Predictors: (Constant), high, medium

Table 34. Model One Summary - Mathematics.

Table 34 displays the summary data for the first linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.274, equates to approximately 27% of the variance in PARCC math scores explained by a student's level of English language proficiency. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.274 represents a large effect size, indicating a strong relationship between English language proficiency and PARCC math performance.

Table 33 shows the results for the estimation of the parameters. When examining the unstandardized *B* coefficient data of the multiple linear regression model, the impact of a student's English language proficiency on the PARCC math scale scores can be seen. Both language proficiency variables are statistically significant. Alone in the model, medium and high English language proficiency are both significant with a *p*-value of 0.000. Accordingly, the regression analysis predicts that increasing a student's English language proficiency to the medium range results in an PARCC math score increase of 30 points, while moving a student to high English language proficiency results in an increase of 54 points. Based on PARCC's scaled scoring system, an increase beyond the low range of English proficiency is equivalent to approximately one performance level increase for grade level math proficiency.

Model two included the four ACCESS assessed language domains of speaking, listening, reading, and writing only. The PARCC scale scores for mathematics were the dependent

variable, with the scaled scores for each of the four language domains as the independent variables.

Table 35

Model Two Summary - Mathematics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
2	0.599*	0.359	0.336	20.616	0.359	15.412	4	110	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale

Table 35. Model Two Summary - Mathematics.

Table 35 displays the summary data for the second linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.336, equates to approximately 34% of the variance in PARCC math scores explained by a student's level of English proficiency in the domains of speaking, listening, reading, and writing. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.336 represents a large effect size, indicating a strong relationship between English language proficiency and PARCC math performance.

Table 33 shows the results for the estimation of the parameters. When examining the unstandardized *B* coefficient data of the multiple linear regression model, the impact of a student's English language proficiency on the PARCC math scale scores can be seen. Both the reading and writing variables are statistically significant. Reading had a *p*-value of 0.001, and writing a *p*-value of 0.002. The speaking and listening variables were not statistically significant. These findings align with those of the second research question. In addition, the regression analysis predicts that a one point increase on a student's ACCESS reading subscore will result in a one-third point increase on the PARCC math assessment. It also predicts that a

one point increase on a student's ACCESS writing subscore will result in a quarter-point increase on the PARCC math assessment.

Model three introduced socioeconomic status variables. The PARCC scale scores for mathematics were the dependent variable, with a student's lunch status serving as a proxy for socioeconomic status for the independent variables. ELs were again divided into three levels measuring different degrees of socioeconomic status: free lunch, reduced lunch, and full pay lunch. This variable was operationalized by creating dummy variables for each distinct level. Students with free lunch served as the reference category for the third model.

Table 36

Model Three Summary - Mathematics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
3	.149*	0.022	0.005	25.238	0.022	1.265	2	112	0.286

* Predictors: (Constant), pay, reduced

Table 36. Model Three Summary - Mathematics.

Table 36 displays the summary data for the third linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.005, equates to less than one tenth of one percent of the variance in PARCC math scores explained by a student's socioeconomic status. According to Cohen (1988) a small effect size has an r^2 value of 0.01 ($r = 0.10$) or greater. Using this measure our r^2 value of 0.005 is not statistically significant and reflects no meaningful relationship between a student's socioeconomic status and PARCC math performance.

Table 33 shows the results for the estimation of the parameters. Neither the reduced or full pay variables are statistically significant within the model. According to this third model, with these variables run in isolation, there is not a statistically significant relationship between

socioeconomic status and mathematics proficiency as measured by PARCC for the student population studied. This finding rejects the hypothesis that full pay students will perform better on this standardized assessment than their lower socioeconomic peers.

Model four included race/ethnicity variables only. The PARCC scale scores for mathematics were the dependent variable, with race/ethnicity the independent variables. ELs self-identified into one of five racial/ethnic categories: Asian, Black, Hispanic, White, and Other. This variable was operationalized by creating dummy variables for each distinct category. Students identifying as Black served as the reference category for the fourth model.

Table 37

Model Four Summary - Mathematics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
4	.388*	0.150	0.119	23.738	0.150	4.865	4	110	0.001

* Predictors: (Constant), asian, hispanic, white, other

Table 37. Model Four Summary - Mathematics.

Table 37 displays the summary data for the fourth linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.119, equates to approximately 12% of the variance in PARCC math scores explained by a student's race/ethnicity. According to Cohen (1988) a small effect size has an r^2 value of 0.01 ($r = 0.10$) or greater. Using this measure our r^2 value of 0.119 indicates a small effect size and to some degree a statistically significant relationship between a student's race/ethnicity and their PARCC math performance.

Table 33 shows the results for the estimation of the parameters. When examining the unstandardized *B* coefficient data of the multiple linear regression model, the impact of a student's race/ethnicity on the PARCC math scale scores can be seen. Each of the race/ethnicity variables were statistically significant and negative. The Asian variable was the least statistically

significant with a *p*-value of 0.067. The three other race/ethnicity variables were all significant at the *p*-value less than or equal to 0.05 level. The *p*-value for the Hispanic variable was 0.011, White was 0.001, and Other at 0.050. As noted, the relationship for each of these variables was negative. The regression analysis predicts that Asian students will score approximately 21 points lower, Hispanic students 31 points lower, White students 38 points lower, and students identifying as Other 39 points lower on the PARCC math assessment than students identifying as Black. This finding rejects the hypothesis that White students will perform better on this standardized assessment than other racial/ethnic groups. Furthermore, this finding affirms the opposite for the student population studied.

Model five included English language proficiency variables and the four ACCESS assessed language domains of speaking, listening, reading, and writing only. There exists a degree of overlap within this model, as a weighted combination of the four language domains - 35% reading + 35% writing + 15% listening + 15% speaking - was used to determine the composite scale scores that define the English language proficiency classifications. The PARCC scale scores for mathematics were the dependent variable, with English language proficiency levels and the scaled scores for each of the four language domains as the independent variables.

Table 38

Model Five Summary - Mathematics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
5	.623*	0.388	0.354	20.340	0.388	11.389	6	108	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale, medium, high

Table 38. Model Five Summary - Mathematics.

Table 38 displays the summary data for the fifth linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.354, equates to approximately 35%

of the variance in PARCC math scores explained by a student's level of English language proficiency in reading, writing, speaking, and listening. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.354 represents a large effect size, indicating a significant relationship between these English language proficiency variables and PARCC math performance.

Table 33 shows the results for the estimation of the parameters. With the introduction of these language proficiency variables together, a student's level of English proficiency as medium is no longer statistically significant. While with a *p*-value of 0.044, the high English proficiency variable remains statistically significant within the model. The reading and writing language variables are still statistically significant with *p*-values of 0.020 and 0.023 respectively. This model again underscores the link between a student's English language proficiency, particularly in the language domains of reading and writing, and increased performance on the PARCC math assessment.

Model six introduced socioeconomic variables. The PARCC scale scores for mathematics were the dependent variable, with English language proficiency levels, scaled scores for each of the four ACCESS assessed language domains, and socioeconomic status as the independent variables.

Table 39

Model Six Summary - Mathematics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
6	.652*	0.425	0.382	19.888	0.425	9.806	8	106	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale, medium, high, reduced, pay

Table 39. Model Six Summary - Mathematics.

Table 39 displays the summary data for the sixth linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.382, equates to approximately 38% of the variance in PARCC math scores explained by the combination of these variables. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.382 represents a large effect size, indicating a significant relationship between the independent variables and PARCC math performance.

Table 33 shows the results for the estimation of the parameters. With the introduction of the socioeconomic variables, a student's level of English proficiency as high remains statistically significant with a *p*-value of 0.033. The reading and writing language variables also remain statistically significant with *p*-values of 0.023 and 0.026 respectively. The predictive impact of each variable also remains relatively constant. In this model, the speaking language variable is significant and negative with a *p*-value of 0.069. Accordingly, an increase in one point on the speaking subdomain score is expected to result in a PARCC math score decrease of one-tenth point. Also in this model, the reduced lunch variable is statistically significant with a *p*-value of 0.023. The full pay lunch variable is not statistically significant, though would be negative if it were. The model predicts that students moving from free to reduced lunch status will increase their PARCC math score by approximately 16 points. As a result, socioeconomic status is an impacting factor within the model, though it appears to be limited with a ceiling at full pay lunch status. Model six further rejects the socioeconomic hypothesis that full pay students will perform at higher levels, but does indicate a relationship between poverty and lower levels of performance on the PARCC math assessment.

Model seven introduced race/ethnicity variables. The PARCC scale scores for mathematics were the dependent variable, with English language proficiency levels, scaled scores for each of the four ACCESS assessed language domains, socioeconomic status, and race/ethnicity as the independent variables.

Table 40

Model Seven Summary - Mathematics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
7	.690*	0.476	0.414	19.358	0.476	7.724	12	102	0.000

* Predictors: (Constant), writing_scale, speaking_scale, listening_scale, reading_scale, medium, high, reduced, pay. asian, hispanic, white, other

Table 40. Model Seven Summary - Mathematics.

Table 40 displays the summary data for the seventh and final linear regression model. According to the multiple linear regression model an adjusted r^2 value of 0.414, equates to approximately 41% of the variance in PARCC math scores explained by the combination of these variables. According to Cohen (1988) a large effect size has an r^2 value of 0.25 ($r = 0.50$) or greater. Using this measure our r^2 value of 0.414 represents a large effect size, indicating a significant relationship between the independent variables and PARCC math performance.

Table 33 shows the results for the estimation of the parameters. With the introduction of the racial/ethnic variables, a student's level of English proficiency as high remains statistically significant, though to a lesser degree with a *p*-value of 0.091. With the introduction of multiple variables into later models, English proficiency levels shifted from being significant to a high degree by itself, to less impacting with the other compounding variables. The reading, writing, and to a lesser degree speaking language variables remain statistically significant with *p*-values of 0.040, 0.020, and 0.094 respectively. The predictive impact of each variable also remains relatively constant, 0.189 for reading, 0.180 for writing, and -0.096 for speaking. The reduced

lunch variable continues to be statistically significant with a *p*-value of 0.027, with its predictive impact of 15 points approximately the same. The full pay lunch variable remains not statistically significant. In this model, each racial/ethnic variable is statistically significant and negative. The Asian variable has a *p*-value of 0.013; Hispanic of 0.006, White of 0.006, and Other of 0.020. The predictive impact of each race/ethnicity variable was negative: approximately 23 points for Asian, 28 points for Hispanic, 27 points for White, and 40 points for Other in comparison to the Black students that served as the reference group.

Accordingly, socioeconomic status is an impacting factor within the model, though it appears to be limited with a cap at full pay status. Model seven again rejects the socioeconomic hypothesis that full pay students will perform at higher levels, but continues to indicate a relationship between poverty and decreased performance on the PARCC math assessment. Model seven also rejects the racial/ethnic hypothesis that students identified as White will perform at higher levels on this standardized mathematics assessment. The findings affirm the opposite. The inclusion of race/ethnicity variables within the model makes each a statistically significant factor, and refutes the hypothesis as each group is predicted to perform at a lower level than the Black reference group.

Summary

The overall research question guiding this study was: how does an ELs' second language proficiency influence his/her achievement on federally mandated standardized content-area assessments? Throughout this chapter, the four research questions and seven corresponding hypotheses were utilized as a framework for presenting the findings of this study. *Table 41* provides a summary of the research findings.

Table 41

Summary of Results

	Research Question	Hypothesis	Reject/ Accept	Interpretation
1	What is the relationship between English language proficiency and achievement on grade level content-area standardized assessments in ELA and mathematics?	1. The more advanced the student's level of language proficiency in English, the higher their reading achievement.	Accept	Through utilizing multiple measures (linear correlation, cross tabulation analysis, Chi-Square test of independence) to assess the strength of the relationship, it is evident that a student's composite ACCESS proficiency score positively correlates to their PARCC ELA scale score. Approximately 44% of the variance in scores was explained by these two variables.
		2. The more advanced the student's level of language proficiency in English, the higher their mathematics achievement.	Accept	Through utilizing multiple measures (linear correlation, cross tabulation analysis, Chi-Square test of independence) to assess the strength of the relationship, it is evident that a student's composite ACCESS proficiency score positively correlates to their PARCC math scale score. Approximately 27% of the variance in scores is explained by these two variables, though the correlation was less significant than what was observed for reading.
2	To what extent does language proficiency in reading, writing, speaking, and listening influence achievement on grade level content-area standardized assessments in ELA and mathematics?	3. A student's English language proficiency in reading and writing will impact their reading achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.	Accept	Linear correlation testing showed a stronger relationship between the language domains of reading (0.683) and writing (0.588) to achievement on the PARCC ELA assessment, than of listening (0.460) and speaking (0.370). Though all of the language domains were statistically significant with respect to reading achievement. According to the multiple linear regression model an adjusted r2 value of 0.485, equates to approximately 49% of the variance in PARCC reading scores explained by the four language domains of the ACCESS assessment - listening, speaking, reading, and writing. A one point increase on the ACCESS reading scale score predicted a 0.524 point increase of the PARCC reading scale score. A one point increase on the ACCESS writing scale score predicted a 0.211 point increase of the PARCC reading scale score.
		4. A student's English language proficiency in reading and writing will impact their mathematics achievement on standardized assessments to a significantly greater extent than the speaking and listening proficiencies.	Accept	Linear correlation testing showed a stronger relationship between the language domains of reading (0.544) and writing (0.530) to achievement on the PARCC math assessment, than of listening (0.361) and speaking (0.244). Though all of the language domains were statistically significant with respect to mathematics achievement. According to the multiple linear regression model an adjusted r2 value of 0.336, equates to approximately 34% of the variance in PARCC math scores explained by the four language domains of the ACCESS assessment - listening, speaking, reading, and writing. A one point increase on the ACCESS reading scale score equaled a 0.302 point increase of the PARCC math scale score. A one point increase on the ACCESS writing scale score equaled a 0.234 point increase of the PARCC math scale score.
3	What is the threshold English language proficiency for ELs to meet/exceed standards on federally mandated standardized assessments in ELA and mathematics?	5. According to Illinois school code ELs are defined as proficient at a composite ACCESS score of 4.8 or beyond, and should thus be able to demonstrate proficiency on grade level content-area assessments.	Reject	Based on the student population sampled the threshold language proficiency for ELs to meet standards on both the PARCC ELA and mathematics assessments was a composite ACCESS proficiency level of 5.1. This score equates to the "bridging" level of language proficiency. This score is beyond the 4.8 threshold the state uses to reclassify ELs.
4	What is the impact of other student demographic factors - ethnicity, socioeconomic status - in relation to English language proficiency for the achievement of ELs on federally mandated standardized assessments in ELA and mathematics?	6. Controlling for other variables, ELs identifying with a race/ethnicity of White will achieve higher levels of reading and mathematics proficiency.	Reject	The inclusion of race/ethnicity variables within the seven linear regression models made each a statistically significant factor, and refuted the hypothesis as each group was predicted to perform at a lower level than the Black reference group.
		7. Controlling for other variables, ELs with full pay lunch status will achieve higher levels of reading and mathematics proficiency.	Reject	The linear regression models indicate that socioeconomic status is an impacting factor when considered in combination with the other language proficiency variables, though it appears to be limited with a ceiling at full pay lunch status. These findings reject the socioeconomic hypothesis that full pay students will perform at higher levels, but do indicate a relationship between poverty and lower levels of performance on the PARCC ELA and math assessments.

Table 41. Summary of Results.

Research Question One

The first research question explored the relationship between English language proficiency and achievement on grade level content-area standardized assessments in ELA and mathematics. Linear correlation was used as a statistical test to determine the strength and direction of the relationship between the two study variables. The Chi-Square test of independence was then used to detect patterns that may indicate relatedness between the two variables.

The first hypothesis stated the more advanced the student's level of language proficiency in English, the higher their reading achievement. Through utilizing multiple methods to assess the strength of the relationship, it is evident that a student's composite ACCESS proficiency score positively correlates to their PARCC ELA scale score. With a Pearson Coefficient, *r*, of 0.661 the strength of the relationship as measured by r^2 was 0.437. Thus, approximately 44% of the variance in scores was explained by these two variables. Furthermore, the Chi-Square value was significant at 39.852, because our calculated value (39.852) exceeded the critical value (15.507), we concluded that there exists a statistically significant difference amongst the reading achievement of EL students based on their level of English proficiency at the 0.05 level. Thus, English language proficiency is clearly a relational factor of reading achievement for EL students on the PARCC ELA assessment, a federally mandated performance-based standardized assessment.

The second hypothesis stated the more advanced the student's level of language proficiency in English, the higher their mathematics achievement. Again utilizing multiple methods to assess the strength of the relationship, it is evident that a student's ACCESS composite proficiency score positively correlates to their PARCC math scale score. With a Pearson Coefficient, *r*, of 0.526 the strength of the relationship as measured by r^2 is 0.277. Approximately 27% of the variance in scores is explained by these two variables. The Chi-Square value is also significant at 33.391, because our calculated value (33.391) exceeded the critical value (15.507), we concluded that there is a statistically significant difference amongst the mathematics achievement of EL students based on their level of English proficiency at the 0.05 level. Hence, English language proficiency is clearly a relational factor of reading

achievement for EL students on the PARCC mathematics assessment. Though a less significant correlation than was seen for reading, this empirical understanding reflects the role of language proficiency as a key indicator of academic achievement for ELs on the PARCC standardized assessment for mathematics.

Based upon linear correlation and the Chi-Square test of independence it is evident that both the first and second hypotheses are supported by the findings in analyzing the data. The evidence from this student population indicates that the more advanced a student's level of language proficiency in English, the higher their reading and mathematics achievement as measured by the PARCC assessment.

Research Question Two

The second research question investigated the relationship between the language proficiencies of reading, writing, speaking, and listening and achievement on grade level content-area standardized assessments in ELA and mathematics. The hypotheses stated that the language proficiencies of reading and writing would impact reading and mathematics achievement to a significantly greater extent than the speaking and listening proficiencies. Linear correlation and multiple linear regression were the statistical tests to determine the strength and direction of the relationship between the study variables.

The linear correlation test showed a stronger relationship between language proficiency in reading (0.683) and writing (0.588) to achievement on the PARCC ELA assessment, than of listening (0.460) and speaking (0.370). Though all of the language proficiencies were statistically significant with respect to reading achievement. According to the multiple linear regression model an adjusted r^2 value of 0.485, equates to approximately 49% of the variance in

PARCC reading scores explained by the four language proficiencies of the ACCESS assessment - listening, speaking, reading, and writing.

When examining the coefficient data of the multiple linear regression model, the impact of each ACCESS language proficiency on the PARCC reading scale scores could be seen. The ACCESS reading scale score, with a *p*-value of 0.000, was statistically significant. The ACCESS writing scale score was also significant with a *p*-value of 0.011. Both the ACCESS listening and speaking scale scores were not significant with *p*-values of 0.335 and 0.731 respectively. Accordingly, utilizing the unstandardized *B* coefficients a one point increase on the ACCESS reading scale score predicted a 0.524 point increase of the PARCC reading scale score. A one point increase on the ACCESS writing scale score predicted a 0.211 point increase of the PARCC reading scale score.

The linear correlation test showed a stronger relationship between language proficiencies in reading (0.544) and writing (0.530) to achievement on the PARCC math assessment, than of listening (0.361) and speaking (0.244). Though all of the language proficiencies were statistically significant with respect to mathematics achievement. According to the multiple linear regression model an adjusted r^2 value of 0.336, equates to approximately 34% of the variance in PARCC math scores explained by the four language proficiencies of the ACCESS assessment - listening, speaking, reading, and writing.

When examining the coefficient data of the multiple linear regression model, the impact of each ACCESS language proficiency on the PARCC math scale scores could be seen. The ACCESS reading scale score, with a *p*-value of 0.001, was statistically significant. The ACCESS writing scale score was also significant with a *p*-value of 0.002. Both the ACCESS

listening and speaking scale scores were not significant with *p*-values of 0.556 and 0.242 respectively. Accordingly, utilizing the unstandardized *B* coefficients a one point increase on the ACCESS reading scale score equaled a 0.302 point increase of the PARCC math scale score. A one point increase on the ACCESS writing scale score equaled a 0.234 point increase of the PARCC math scale score.

Using both correlation and multiple linear regression as statistical tests, both hypotheses were supported in that a student's English language proficiency in reading and writing impacted their reading and mathematics achievement on standardized assessment to a significantly greater extent than their speaking and listening proficiencies.

Research Question Three

The third research question examined the threshold English language proficiency for ELs to meet/exceed standards on federally mandated standardized assessments in ELA and mathematics. The hypothesis pertaining to this question was that a composite ACCESS score of 4.8 or beyond, where Illinois school code defines ELs as proficient, should demonstrate proficiency on grade level content-area assessments. Mean and median ACCESS composite scores were compared within the ranges of PARCC academic proficiency levels, which allowed for determining the overall composite ACCESS score that serves as the threshold for achieving proficiency on the PARCC ELA and mathematics assessment for our particular student population. Histograms and cross tabulation tables were used to visually illustrate the data.

Based on the student population sampled the threshold language proficiency for ELs to meet standards on both the PARCC ELA and mathematics assessments was a composite ACCESS proficiency level of 5.1. This score equates to the "bridging" level of language

proficiency. This score is beyond the 4.8 threshold the state uses to reclassify ELs and demonstrates a misalignment between research, policy, and current assessment practices.

Research Question Four

The fourth research question considered the impact of other student demographic factors - ethnicity, socioeconomic status - in relation to English language proficiency for the achievement of ELs on federally mandated standardized assessments in ELA and mathematics. The first hypothesis for this research question was that by controlling for other variables, ELs identifying with a race/ethnicity of White would achieve higher levels of reading and mathematics proficiency. The second hypothesis for this question was that controlling for other variables, ELs with full pay lunch status would achieve higher levels of reading and mathematics proficiency. Seven linear regression models were used to answer this final research question and test its two accompanying hypotheses.

Model one included English language proficiency variables only and found that both language proficiency variables were statistically significant for both reading and math achievement. Alone in the model, medium and high English language proficiency had *p*-values of 0.000 for both content area assessments. Accordingly, the regression analysis predicted that increasing a student's English language proficiency to the medium range resulted in an PARCC score increase of 44 points for ELA and 30 points for math, while moving a student to high English language proficiency results in an increase of 79 points for ELA and 54 points for math. Based on these results, hypotheses one and two for the first research questions were again supported.

Model two included the four ACCESS assessed language domains of speaking, listening, reading, and writing only. In alignment with the findings of the second research question, both the reading and writing variables were statistically significant for both the ELA and math PARCC assessments. Reading had a *p*-value of 0.000 for ELA and 0.001 for math. Writing had a *p*-value of 0.011 for ELA and 0.002 for math. The speaking and listening variables were not statistically significant for either assessment in this model. The regression analysis predicted an increase in both the PARCC ELA and math assessment with a one point increase on a student's ACCESS reading and writing subscores. Based on these results, hypotheses three and four for the second research questions were again supported.

Model three introduced socioeconomic variables. Neither the reduced or full pay variables were statistically significant within the model. According to model three there was not a statistically significant relationship between socioeconomic status and reading/mathematics proficiency as measured by PARCC for the student population studied. These findings rejected the seventh hypothesis that full pay students will perform better on this standardized assessment than their lower socioeconomic peers.

Model four included race/ethnicity variables only. With respect to reading, the White variable was the only one of significance with a *p*-value of 0.017. Accordingly, the regression analysis predicted that White students will score approximately 35 points lower on the PARCC ELA assessment in comparison to the Black reference group. For math, each of the racial/ethnic variables was statistically significant and negative. The Asian variable was the least statistically significant with a *p*-value of 0.067. The *p*-value for the Hispanic variable was 0.011, White was 0.001, and Other at 0.050. As noted, the relationship for each of these variables was negative.

The regression analysis predicted that all of the racial/ethnic groups would perform lower in relation to the Black reference group: approximately 21 points for Asian students, 31 points for Hispanic students 31, 38 points for White students, and 39 points for students identifying as Other. These findings rejected the sixth hypothesis that White students would perform better on this standardized assessment than other racial/ethnic groups. Furthermore, these findings affirmed the opposite for the student population studied.

Model five included English language proficiency variables and the four ACCESS assessed language domains of speaking, listening, reading, and writing only. Regarding reading proficiency, with the introduction of these language proficiency variables together, a student's level of English proficiency as medium or high was no longer statistically significant. The reading and writing language variables were still statistically significant with *p*-values of 0.000 and 0.027 respectively. In relation to math proficiency, the high English proficiency variable remained statistically significant with a *p*-value of 0.044. The reading and writing language variables were still statistically significant with *p*-values of 0.020 and 0.023 respectively. Model five further reaffirmed hypotheses three and four of research question two, and more clearly establishes the connection between a student's English language proficiency, particularly in the language domains of reading and writing, and increased performance on the PARCC ELA and math assessments.

Model six introduced socioeconomic variables. In relation to reading proficiency, the language domains of reading and writing remained constant and significant, while reduced lunch status was also significant with a *p*-value of 0.026. The model predicted that students moving from free to reduced lunch status would increase their PARCC ELA score by approximately 17

points. Similarly for mathematics proficiency, the language domains of reading and writing remained constant and significant, while reduced lunch status was also significant with a *p*-value of 0.023. High language proficiency and the language domain of speaking were also significant within this model. Model six predicted that students moving from free to reduced lunch status would increase their PARCC math score by approximately 16 points. Model six demonstrated that socioeconomic status is an impacting factor when considered in combination with the other language proficiency variables, though it appears to be limited with a ceiling at full pay lunch status. These findings further reject the socioeconomic hypothesis that full pay students will perform at higher levels, but do indicate a relationship between poverty and lower levels of performance on the PARCC ELA and math assessments.

Model seven introduced race/ethnicity variables. In reference to reading proficiency, the language domains of reading and writing and reduced lunch status remained constant and significant. All of the racial/ethnic variables were significant and negative. The Asian variable had a *p*-value of 0.043; Hispanic of 0.049, White of 0.018, and Other of 0.010. The predictive impact of each race/ethnicity variable was negative: approximately 21 points for Asian, 23 points for Hispanic, 26 points for White, and 32 points for Other in comparison to the Black students that served as the reference group. In terms of mathematics, high language proficiency, the domains of reading, writing, and speaking and reduced lunch status remained constant and significant. Like reading, all of the racial/ethnic variables were significant and negative. The Asian variable has a *p*-value of 0.013; Hispanic of 0.006, White of 0.006, and Other of 0.020. The predictive impact of each race/ethnicity variable was negative: approximately 23 points for Asian, 28 points for Hispanic, 27 points for White, and 40 points for Other in comparison to the

Black students that served as the reference group. Model seven again rejected the seventh hypothesis that full pay students will perform at higher levels, but continued to indicate a relationship between poverty and decreased performance on the PARCC math assessment. Model seven also rejected the sixth hypothesis that students identified as White would perform at higher levels on this standardized mathematics assessment. The findings affirmed the opposite. The inclusion of race/ethnicity variables within the model made each a statistically significant factor, and refuted the hypothesis as each group was predicted to perform at a lower level than the Black reference group.

Conclusion

This chapter provided an overview of the data and analysis of the results relating to each of the research questions and underlying hypotheses. The purpose of this quasi-experimental correlational design quantitative study was to examine the relationship between ELs' second language proficiency and achievement on federally mandated standardized content-area assessments. In summary, these findings affirmed hypotheses one and two for the first research question. Hypotheses three and four with respect to the second research question were also supported by the test results. The fifth hypothesis, related to research question three, was rejected based on the data. Additionally, the results of the seven linear regression models led to hypotheses six and seven for the fourth research question being rejected.

Across all four of the research questions, the findings indicate a clear connection between English language proficiency and performance on content-area standardized assessments for ELA and mathematics for the EL population studied. It is hoped that a more clearly developed empirical understanding of the relationship between English language proficiency and grade

level content-area achievement, as measured by performance-based standardized tests, will help to inform both legislators and educators as to assessment practices and policies that are most equitable and valid for ELs and their school communities.

CHAPTER FIVE. CONCLUSION.

Summary

Throughout the history of the United States, literacy tests have been utilized as a means to exclude and marginalize immigrant populations, and in the absence of a national language policy, have served as de facto language policy (Crawford, 2000; Menken, 2008; Wiley & Wright, 2004).

Throughout the twenty-first century, ELs have been among the fastest-growing populations in our nation's schools, comprising ten percent of the student population nationwide (United States Department of Education, 2017). ELs comprise a heterogeneous group of students with different cultural and linguistic backgrounds (Center for Applied Linguistics, 2017; Migration Policy Institute, 2018; United States Department of Education, 2017). As measured through school demographics, the United States is more diverse than ever before. Public school enrollment in 2017 shows that, for the first time ever, majority-minority children constituted 51% of the student population, and students from low income households also made up 51% of the population (Center for Applied Linguistics, 2018). According to the Illinois State Report Card during the 2018-19 school year, 12.1% of students within the state's K-12 system were classified as ELs. This constitutes over 240,000 children, reflecting a gradual increase of over 3% for the state's EL population in the past decade (Illinois State Report Card, 2019).

Federal education legislation of the past two decades has demonstrated a significant shift towards emphasizing standardized assessments within a system of high-stakes accountability (Abedi, 2003; Menken, 2009; Hopkins, Thompson, Linquanti, Hakuta, & August, 2013). While these policies have impacted all students, there has been a broader and disproportionately greater

negative impact on ELs and the school communities that serve them (Ho, 2008; Kortez, 2008; Linn, 2005, 2008). The *Every Student Succeeds Act* (ESSA) mandates that ELs must take both academic content-area and language proficiency assessments (Lyons & Dadey, 2017). The rationale for including ELs in these mandated English language proficiency and grade level content-area assessments and overarching systems of accountability is to ensure their educational progress relative to English language development and grade level content-area standards (Abedi, 2003; Menken, 2009). However, the most salient factor concerning the validity of assessment for ELs is that these standardized assessments are not designed, nor implemented, with the EL student population in mind (Abedi, 2011; Kopriva, 2000; Solano-Flores, 2011; Trumbull et al, 2011).

Research findings concerning the assessment of ELs highlight a significant difference in the academic performance of these students from that of their native English speaking classmates (Abedi, 2004, 2006; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Solano-Flores & Trumbull, 2003). ELs face a significant hurdle in learning both English and academic content in English at the same time. Research indicates that it takes five to seven years for most ELs to gain mastery of academic English at a level comparable to native English speaking peers (Collier & Thomas 1989, 2005, 2009; Cummins, 1981, 2012; Hakuta, Butler, Witt, 2000). This challenge is intensified on grade level content-area assessments that are densely worded with academic vocabulary and ripe with linguistic complexity (Abedi, 2002, 2006, 2011; Solano-Flores, 2008). Additionally, ELs are often supported on these assessments with ineffective accommodations that were not designed to meet their specific linguistic needs (Abedi, 2007, 2011; Abedi, Hofstetter, & Lord, 2004; Lara & Chia, 2011). Hence, there exists

substantial misalignment between the research, aims, policies, and practices surrounding the assessment and educational experiences of ELs.

The growing population of ELs within the United States operating within a system of standardized assessment and high-stakes accountability, leads to an increased sense of urgency in addressing these issues, remediating achievement gaps, and better meeting the educational needs of this unique student population. There exists a need to further address the contradiction between the stated aims and actual impact of assessments at the federal, state, district, school, and classroom levels. Investigating current assessment policies and practices, while seeking to create a more valid and equitable system of assessment for ELs, is important as these tests currently drive the instructional practices, decision-making, and educational opportunities of ELs in the United States (Menken, 2009).

The purpose of this quasi-experimental correlational design quantitative study was to examine the relationship between ELs' second language proficiency and achievement on federally mandated standardized content-area assessments. This study sought to address the overarching issues concerning the processes of second language development and assessment of ELs within the context of public education in the United States. The overall research question guiding this study was: how does an ELs' second language proficiency influence his/her achievement on federally mandated standardized content-area assessments? The four research questions and seven corresponding hypotheses are utilized in the following section as a framework for discussing the findings of this study. The results of this study seek to inform policymakers and educators at the state, district, school, and classroom level as to the most

equitable, valid, and just assessment practices and policies for ELs and the school communities that serve them.

Findings

In this section, the findings of the study are discussed and related to each of the research questions and accompanying hypotheses, as well as, linked to theories and research that contribute to an overarching theoretical framework.

Research question 1. What is the relationship between English language proficiency and achievement on grade level content-area standardized assessments in ELA and mathematics?

Hypothesis 1. The more advanced the student's level of language proficiency in English, the higher their reading achievement.

Hypothesis 2. The more advanced the student's level of language proficiency in English, the higher their mathematics achievement.

Based upon the results of linear correlation and the Chi-Square test of independence it is evident that a student's composite ACCESS proficiency score positively correlates to their PARCC ELA and mathematics scale score. These findings reinforce the notion that English language proficiency is clearly a relational factor of reading and math achievement for EL students on the PARCC assessment, a federally mandated performance-based standardized assessment. These conclusions are consistent with the prevailing body of research that finds the burden of complex language on assessments is a significant factor that contributes to the gap in performance between ELs and their non-EL peers (Abedi, 2006; Abedi, Hofstetter, & Lord, 2004; Solano-Flores & Trumbull, 2003).

Time is clearly an important consideration of these findings. Given a variety of factors, the amount of time that it takes for an EL to acquire academic language proficiency is between five and seven years (Collier & Thomas 1989, 2005, 2009; Cummins, 1981, 2012; Hakuta, Butler, Witt, 2000). English proficiency encompasses not just the oral proficiency of students, but a child's reading, writing, speaking, and listening proficiencies as measured by grade level standardized assessments across various content areas (Collier & Thomas, 1995, 2007, 2009; Cook, 2012; Cummins, 1981; Hakuta, Butler, Witt, 2000). Conversational fluency however is not the benchmark by which ELs are measured and held accountable in comparison to their native monolingual peers. Academic English is the language of school, the language of testing, and also the language in which academic competence is assessed. On the PARCC assessment and other similar performance-based standardized measures, Academic English is the language of achievement and thus ultimately success (Collier & Thomas, 2009; Cook, 2012; Cummins, 2012; Hakuta, Butler, Witt, 2000). It is not just receptive language skills of students that are important for demonstrating mastery towards grade level standards on these assessments, but productive language skills that require additional time to develop. This holds particularly true for next generation performance-based assessments, like PARCC, that require students to engage with, comprehend, and produce multiple literacies to demonstrate their content-area knowledge. Linguistic demands are inherent in standardized assessments taken by ELs as there is a presumed level of English language proficiency in the construct of these tests that are not related to the content knowledge being assessed (Trumbull & Solano-Flores, 2011). The overarching theme of EL performance on standardized assessments being significantly impacted by language proficiency and linguistic complexity rather than content knowledge is paramount across the

research (Abedi, 2002, 2003, 2009, 2011; Solano-Flores, 2008; Trumbull & Solano-Flores, 2011). The evidence from the results in connection with the first research question empirically supports this connection between English language proficiency and grade level achievement.

The most universal factor impacting the assessment of ELs is that standardized assessments are not designed, nor implemented, with the EL student population in mind (Abedi, 2011; Kopriva, 2000; Solano-Flores, 2011; Trumbull et al, 2011). Increased linguistic demand decreases the validity and reliability of assessments for ELs (Abedi, 2002, 2006; Solano-Flores, 2008). For ELs the greatest source of measurement error, and thus reduced assessment validity, comes from unnecessary linguistic complexity (Abedi, 2002, 2006; Solano-Flores, 2008). Unnecessarily dense language serves to distort the construct of individual questions and thus the entirety of assessments for ELs, measuring their language proficiency rather than the intended content objectives (Abedi, 2011). When language demands are lessened the performance gaps between ELs and non-ELs begin to close (Abedi, 2009). Additional research further supports that language factors have a more significant impact on ELs than on native English speakers (Abedi, 2006; Abedi et al, 2003). Abedi (2003) linked the impact of this linguistic demand with the performance of ELs on standardized achievement tests and concluded that greater language usage expanded the performance gap between ELs and their English speaking peers. Similarly, the results from the linear correlation and Chi-Square tests of this study further support these findings. The significant and positive correlation of ACCESS scaled scores and PARCC scaled scores demonstrate a clear link between English language proficiency and reading and math achievement. It can reasonably be inferred that the effects would be even greater for students whose native language was English and grew up within the mainstream culture.

Research question 2. To what extent does English language proficiency in reading, writing, speaking, and listening influence achievement on grade level content-area standardized assessments in ELA and mathematics?

Hypothesis 3. A student's English language proficiency in reading and writing will impact their reading achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.

Hypothesis 4. A student's English language proficiency in reading and writing will impact their mathematics achievement on standardized assessments to a significantly greater extent than their speaking and listening proficiencies.

Based upon the results of correlation testing and multiple linear regression it is evident that English language proficiency in reading and writing impact a student's reading and mathematics achievement on the PARCC standardized assessment to a significantly greater extent than their speaking and listening proficiencies. "Research shows that ELs develop receptive skills more rapidly than the productive skills needed to write essays in academic English (Cummins, 1992; García & Menken, 2006)" (Menken, 2008, p. 182). The findings of this study reinforce the notion that it is not just receptive language skills of students that are important for demonstrating mastery towards grade level standards on performance-based assessments, like PARCC, but also productive language skills that require additional time to develop.

Returning to the research regarding length of time, processes, and stages of second language acquisition, Cummins' (1981) language quadrant framework and iceberg analogy represent continuums of contextual support and communicative competence. Cummins' (1981)

delineates a distinction between highly contextual surface language at low levels of cognitive demand and deeper context-reduced language at higher levels of cognitive demand. The former, cognitively undemanding and within context, develop in the earlier stages of linguistic acquisition, whereas abstract cognitively demanding language develops later. By their very nature, standardized assessments lie outside the typical context of classroom experiences, lack meaningful linguistic support, and feature a disconnected series of questions/tasks unrelated to the lived experience of most students. Standardized assessments are the greatest educational examples of context reduced and cognitively demanding activities (Cummins, 1981,1989, 2000). While this may pose a challenge for many students, it serves as a double hurdle for ELs who require a context rich environment in order for learning to be comprehensible and meaningful.

Furthermore, conversational ability is often misperceived by educators as a sign of fluency, while deeper, more submerged levels of communication proficiency lie below the surface not yet fully developed. ELs often present as more fluent than they really are, requiring additional time - five to seven years - to develop more advanced levels of fluency (Collier, 1987, 1989; Collier & Thomas, 2007, 2009; Cummins 1981, 2012; Krashen & Terrell, 1983). Derived from Krashen and Terrell's (1983) stages of second language acquisition, the WIDA performance definitions describe the significant markers of language development, representing a continuum of language learning rather than a leveled linear progression. This framework is the second language resource with the greatest impact upon most educational systems, particularly in Illinois and the other 39 states within the WIDA consortium. These understandings, and the blossoming research around translanguaging, shaped the hypothesis that the more evolved

language proficiencies of reading and writing would have a greater impact upon the standardized assessment performance of ELs than speaking and listening.

Further building upon this idea, Abedi et al (2003) found that assessment outcomes for ELs are not structurally as consistent as those of non ELs due to linguistic demand as a source of construct irrelevance. Further research has identified specific sources of language demands - vocabulary, syntax, and discourse - that create issues for ELs and other students by conflating knowledge and skills with language proficiency (Abedi, Hofstetter, Baker, & Lord, 2001; Trumbull & Solano-Flores, 2011). With respect to vocabulary, false cognates, unfamiliar words and phrases, and lengthy words all lead to increased linguistic demand. Syntactically, unfamiliar tenses, negation, prepositional phrases, and compound sentences are amongst several grammatical considerations that increase the linguistic requirement of an assessment. These challenges for ELs are especially true in standardized assessments like PARCC, that feature extensive grade level texts on a variety of literary, scientific, and historical topics. Lengthy problem statements, multiple instructional steps, and passive voice all contribute to the construction of assessment discourse that add an extraneous language burden for ELs (Abedi, Hofstetter, Baker, & Lord, 2001; Trumbull & Solano-Flores, 2011). All of this describes the PARCC testing experience in both ELA and mathematics for ELs. Though characterized as content-area assessments for grade level reading and math standards, these assessments embed language in complex ways that distort the construct in ways unique for ELs and end up testing constructs that were never intended. These linguistic demands, independent of other cultural demands, add to the complexity of assessment for ELs and stand as a barrier for gauging their true depth of knowledge and understanding. These points were further supported by the findings

throughout this study. More specifically, the outcomes of correlation testing and multiple linear regression in relation to the second research question permit us to conclude that English language proficiency in reading and writing impact a student's reading and mathematics achievement on standardized assessment to a significantly greater extent than their speaking and listening proficiencies. This reinforces the importance of productive language skills that require time, five to seven years for ELs to develop with native like proficiency, to be able to accurately demonstrate their content knowledge on grade level performance-based standardized assessments.

Research question 3. What is the threshold English language proficiency for ELs to meet/exceed standards on federally mandated standardized assessments in ELA and mathematics?

Hypothesis 5. According to Illinois school code ELs are defined as proficient at a composite ACCESS score of 4.8 or beyond, and should thus be able to demonstrate proficiency on grade level content-area assessments.

Based upon the comparison of mean and median ACCESS composite scores for the student population studied within the ranges of PARCC academic proficiency levels, the threshold language proficiency to meet standards on both the PARCC ELA and mathematics assessments was a language proficiency level of 5.1.

ELs at a level 5.1 are classified as "bridging" within the WIDA framework. According to the *WIDA Performance Definitions for the Levels of English Language Proficiency in Grades K-12,* third through fifth grade ELs at the bridging level will be able to process, understand, produce, and use: "specialized or technical language reflective of the content areas at grade level;

a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse as required by the specified grade level; oral or written communication in English comparable to English-proficient peers" (WIDA, 2012). The State of Illinois reclassifies ELs as no longer needing targeted language support at level 4.8, "expanding." The findings of this study conclude that a minimum language proficiency level of 5.1 is needed to meet grade level standards at a rate comparable with English speaking peers. *Table 41* outlines the WIDA CAN DO descriptors, the language ELs can process and produce, for third through fifth grade students at both a level 4 (expanding) and level 5 (bridging).

Table 42

WIDA CAN DO Descriptors: Grade Level Cluster 3-5

	Level 4 Expanding	**Level 5 Bridging**
Listening	- Interpret oral information and apply to new situation - Identify illustrated main ideas and supporting details from oral discourse - Infer from and act on oral information - Role play the work of authors, mathematicians, scientists, historians from oral reading, video, or multi-media	- Carry out oral instructionals containing grade-level content-based language - Construct models or use manipulatives to problem-solve based on oral discourse - Distinguish between literal and figurative language in oral discourse - Form opinions of people, places, or ideas from oral scenarios
Speaking	- Answer opinion questions with supporting details - Discuss stories, issues, and concepts - Give content-based oral reports - Offer creative solutions to issues/problems - Compare/contrast content-based functions and relationships	- Justify/defend opinions or explanations with evidence - Give content-based presentations using technical vocabulary - Sequence steps in grade-level problem-solving - Explain in detail results of inquiry (e.g., scientific experiments)
Reading	- Classify features of various genres of text (e.g. "and they lived happily ever after"-fairy tales) - Match graphic organizers to different texts (e.g., compare/contrast with Venn diagram) - Find details that support main ideas - Differentiate between fact and opinion narrative and expository text	- Summarize information from multiple related sources - Answer analytical questions about grade-level text - Identify, explain, and give examples of figures of speech - Draw conclusions from explicit and implicit text at or near grade level
Writing	- Take notes using graphic organizers - Summarize content-based information - Author multiple forms of writing (e.g., expository, narrative, persuasive) from models - Explain strategies or use of information in solving problems	- Produce extended responses of original text approaching grade level - Apply content-based information to new contexts - Connect or integrate personal experiences with literature/content - Create grade-level stories or reports

Table 42. WIDA CAN DO Descriptors: Grade Level Cluster 3-5.

One of the key differences between these two levels of language proficiency is the ability to create and apply complex language. As previously discussed, these productive linguistic skills are fundamental to success on performance-based standardized assessments. The difference in expectation with respect to these skills is evident in the performance descriptors (i.e. "summarize content-based information" versus "apply context-based information to new contexts"). Furthermore, according to the performance descriptors, ELs through a level four, including where the state currently draws its cutoff, require visual, graphic, or interactive support to be able to achieve these descriptors. Not until level 5 do the WIDA descriptors state that a student should be able to work with a text "at or near grade level" or that they should "create grade-level stories or reports." Thus when being assessed on grade level text beyond what can reasonably be expected based on the time required for the development of academic language in a second language, the issue of ELs responding to the question as intended, even with mastery of the content and skills, is a significant factor limiting the performance, validity, and reliability of standardized assessments for ELs (Abedi, 2002, 2006; Durán, 2011). ELs must have adequate linguistic and cultural knowledge to be able to navigate the assessment and understand the demands being placed upon them (Durán, 2011; Solano-Flores & Nelson-Barber, 2001; Solano-Flores, 2011). Hence, there exists a clear imbalance between the framework that serves as the state's foundation for educational systems working with ELs and its own assessment policies and practices.

This in addition to the primary concern of standardized assessments, that the construct and design processes of these tests are not inclusive of language factors and do not fully consider the multifaceted linguistic needs of ELs (Abedi, 2011; Kopriva, 2000; Rivera & Vincent, 1997).

While these assessments may be constructed with universally accepted design principles and may indeed be valid and reliable for native English speakers, this is not necessarily so for ELs (Abedi, 2011). This is again evident and underscored by the findings of this study, specifically highlighted by examining the increase of mean composite ACCESS scaled score for each increasing range of PARCC performance level. It is beyond question that for the student population of this study, the higher a child's level of English language proficiency, the greater the probability of meeting grade level standards on the PARCC standardized assessment. The findings of this study in relation to the third research question further critique this disconnect and provide empirical evidence for a misalignment of practice and policy.

Research question 4. What is the impact of other student demographic factors - race/ethnicity, socioeconomic status - in relation to English language proficiency for the achievement of ELs on federally mandated standardized assessments in ELA and mathematics?

Hypothesis 6. Controlling for other variables, ELs identifying with a race/ethnicity of White will achieve higher levels of reading and mathematics proficiency.

Hypothesis 7. Controlling for other variables, ELs with full pay lunch status will achieve higher levels of reading and mathematics proficiency.

The results of the fourth research question served to reinforce the earlier findings of this study. Based upon the results of seven linear regression models, the findings from earlier in the study regarding the impact of language proficiency on content-area achievement and added significance of reading and writing proficiency were upheld. Accordingly, these findings

affirmed the first four hypotheses addressing the first and second research questions in this study. Similarly, the results of the regression models rejected the fifth hypothesis, of grade level readiness at a language proficiency level of 4.8, for all of the reasons outlined in the previous section.

As has been detailed throughout this study and other research, ELs are an incredibly diverse population, and the reliability, validity, and effectiveness of any assessment depends upon meeting the unique needs of each particular child (Abedi, 2011; Center for Applied Linguistics, 2017; Kopriva, 2000; Migration Policy Institute, 2018; Solano-Flores, 2011). Consideration should be given to a range of demographic variables - English language proficiency, time in the United States, reading proficiency, level of schooling in home language, as well as, other factors that statistically impact standardized assessment scores, socioeconomic status, parent educational levels, etc. (Abedi et al, 2004). Further expanding upon these findings, Menken (2008) stated, "it is of concern that test scores are closely aligned to race, class, and English proficiency level, yet serve as the indicator for accountability measures. As the system is now, we are simply perpetuating the inequities of students when they enter school, through the inequities of their test scores when they leave" (p. 187). Building upon this perspective, the seven regression models sought to examine the influence of demographic factors - ethnicity, socioeconomic status - in relation to English language proficiency for the achievement of the EL population on the PARCC ELA and mathematics assessments.

The regression models supported a significant, but limited relationship concerning socioeconomic status and PARCC achievement when considered with the other race/ethnicity and language proficiency variables. The results found that socioeconomic status did impact

reading and math performance, but that the effects were limited with an upper limit at full pay lunch status. For the student population of this study, there was a pronounced difference in performance between those living in poverty and others slightly above it. Additional financial resources beyond that threshold did not translate to higher test scores in this study. As will be discussed further below, these findings should be expanded to include a larger sample size. It is not unreasonable to conclude based on these findings that socioeconomic status is clearly a factor in student achievement, though it may play a less significant role relative to English language proficiency in the assessment performance of ELs.

The regression models supported a significant relationship between race/ethnicity and PARCC achievement when considered with the socioeconomic and language proficiency variables. The inclusion of racial/ethnic variables within the model made each a statistically significant factor. Though the hypothesis was refuted as each group was predicted to perform at a lower level than the Black reference group. This hypothesis was shaped by Solano-Flores and Nelson-Barber's (2001) theory of cultural validity, that "the effectiveness with which [...] assessment addresses the socio-cultural influences that shape student thinking and the ways in which students make sense of [...] items and respond to them" (p.555). It was expected that ELs identifying as White and other "high achieving" racial/ethnic groups would perform at higher levels on the PARCC assessment, as they would be more apt to see their "set of values, beliefs, experiences, communication patterns, teaching and learning styles, and epistemologies" within the text selections and problem sets of the assessment. These results, while important, are limited and warrant further investigation. The sample size of Black students for this study was small (n=5). Accordingly, the performance of these students had a significant impact upon the

study as they served as the reference group within the regression models. In addition to this, the White student group was not representative of a group of students accessing white privilege, steeped in its language and culture. As was previously discussed, the district featured in this study serves a large Middle Eastern population, with many self-identifying as White on demographic forms. The generalizability of the racial and ethnic identities for the student population is a clear limitation of the study. The findings, though different than expected, do further support the body of research in identifying race and ethnicity as demographic factors that do significantly impact student achievement on standardized assessment. Accordingly, the results of this study support the need for assessments to be developed in a culturally valid and responsive manner, with the consideration of all students taken into account at the beginning of the design and development process.

The results of the seven regression models supported the findings across all four of the research questions and indicate a clear connection between English language proficiency and performance on content-area standardized assessments for ELA and mathematics for the EL population studied.

Implications

The results of this study empirically demonstrated that the PARCC assessment for both ELA and mathematics is a linguistically laden challenge for ELs in a way that it is not for students whose primary language is English. The results clearly elucidate the relationship between English language proficiency and achievement on PARCC, a federally mandated performance-based content-area assessment. While these findings regarding the impact of English language proficiency on measuring achievement for ELs are not new (Abedi, 2004,

2006; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Grisson, 2004; Solano-Flores & Trumbull, 2003), this study does further contribute to this body of research and build upon it to encompass next generation performance-based assessments. Furthermore, this study provides specific information and insight regarding the validity and reliability of standardized assessment practices within the State of Illinois at a moment of decision and opportunity.

This study is particularly timely and relevant as the State of Illinois reviews its annual grade level content-area assessment for ELA and mathematics, and considers a shift towards a new assessment for the 2020-21 school year. The Illinois State Board of Education (ISBE) is currently gathering feedback from stakeholders regarding current assessment practices.

> ISBE launched a survey today to gather feedback on state and federally required student assessments...The feedback will help ISBE prioritize short-term improvements and long-term enhancements to the state assessment system. ISBE has identified equity, inclusivity, usefulness, and balance as the central values defining its vision for state assessments (ISBE, 2019).

ISBE has committed to gathering and reviewing this feedback as it begins the process of making recommendations for assessment systems and practices heading into the 2020-21 school year.

Given the ongoing emphasis of standardized assessment as a driver for accountability and school improvement efforts throughout all levels of education, the most profound implication of this research is very simply to share it. To stand up, say something, and do something. Each of us, no matter our formal position, have a voice and stake in the equitable and impactful education of all children. The findings of this, and other studies, are clear. Standardized assessment is not an equitable experience for ELs and yields specious results that are then utilized in

misunderstood, inappropriate, and even unethical ways. Examples include the singular use of these assessments to determine class placements, instructional programming, and academic interventions often separated from their peers. That being said, say something. Make this matter a point of conversation and priority within your sphere of influence. In order to work together to address the issue, we first have to be willing to acknowledge it as a problem. Collectively, we must build greater capacity for understanding and recognizing the impact of current educational and de facto language policies upon our children, schools, and communities. We must acknowledge the important role that all educational actors play in navigating the complex layers of policy implementation, and in negotiating equitable experiences for all children. Based on the findings of this study, the following recommendations are made for policymakers, district/school leaders, and teachers.

Policymakers

Policymakers need to formally recognize that standardized assessments are not a reliable, valid, and meaningful experience for all students. They need to acknowledge that for too long these tests have operated in a system of high-stakes accountability that privileges few at the cost of many. This recognition is necessary for moving forward with school and communities in creating a new order, system, and conception of educational success.

Beyond that, policymakers need to move beyond an over reliance on standardized assessments. The current system in Illinois links seventy percent of a school's rating to standardized assessment scores. This system needs to be shifted from one of accountability to responsibility, and examine the multitude of other measures that can be used to assess student learning. From portfolios featuring student works to classroom, school, and district assessments

there are multiple measures for portraying student learning. Additionally, growth should be used as a measure to monitor individual student progress and hold schools accountable for ensuring the development of each unique child. The current growth measurement utilized by the state is characterized by a competitive model that fails to take into consideration unique student characteristics. In it, the growth of ELs is compared side by side with that of monolingual English speaking peers. Ensuring the growth and academic development of each child need not be a competitive exercise, but is better served by a system that encourages high expectations and meaningful levels of growth for all students.

These changes would help to reframe current narratives regarding the effectiveness of school reform. Moreover, policymakers must strategically redefine a focus that emphasizes the ability of ELs to learn and achieve. Current policies, systems, and rhetoric all fuel deficit-based approaches to the EL student population and community. Institutionally, we must adopt an asset-based and growth-minded perspective with respect to ELs and to the incredible promise they hold for our schools, communities, and state.

Part of this work includes redefining the classification of ELs at a point that is more in line with research, including that of this study. At a language proficiency level of 4.8, Illinois' threshold for reclassifying ELs is lower than most other states. Accordingly, the targeted language supports and financial resources that accompany such programming are no longer required to be provided to students once they are no longer classified as ELs. Illinois should follow the lead of other states in adopting a proficiency level of 6.0 to more fully support resources, programming, and instruction for ELs. Furthermore, the state must play a more active role in developing teachers, both pre-service and in the field, to better meet the needs of

culturally and linguistically diverse students. Increased emphasis must be placed on recognizing the role of language within assessment and instruction, as well as, for providing research-based strategies that make content more comprehensible for linguistically diverse learners. Additionally, providing teachers with the resources, research, and time to reflect to ensure their practices and pedagogies are culturally valid and responsive is important and never ending work.

Lastly, it is likely that moving forward large-scale standardized assessment will continue to persist. It is imperative that policymakers use their position, and financial holdings, to push for testing policies that align with best-practice assessment and accommodations research. As Trumbull & Koelsch (2011) note, "achieving validity and equity in the assessment of ELLs requires addressing all of the elements of assessment: content (including language of instructions and any other text), format, administration, scoring criteria and procedures, score interpretation, and use" (Farr & Trumbull, 1997; Solano-Flores et al, 2001; Solano-Flores & Trumbull, 2003; Solano-Flores, Trumbull, & Nelson-Barber, 2002, p. 196). Current research efforts, led by Abedi and others, have expanded the general understanding of assessment accommodations most appropriate to ELs, strongly advocating against a "one size fits all" approach to EL testing accommodations. Additional recommendations include that all students, ELs and non-ELs, benefit from clear assessment language. This is a simple linguistic change that is to the benefit of every student engaging with a standardized assessment. Language can then be further modified for ELs and other student populations by reducing the use of low-frequency vocabulary and complex language structures that are irrelevant to the content being tested. This accommodation is an effective and valid practice as it does not impact the performance of English proficient students (Abedi, Hofstetter, Lord, 2004; Lara & Chia, 2011; Trumbull &

Koelsch, 2011). Personalized dictionaries are identified as an effective support for ELs, that also do not affect the score of English proficient students. Again all students could benefit from such accommodation (Abedi, Hofstter, Lord, 2004). It will be in addressing the totality of these factors, that the assessment process becomes a more valid, equitable, and ultimately meaningful experience for ELs and the school communities that serve them.

Policymakers play a significant role within the current system and have tremendous authority in transforming it. The current moment provides the opportunity to learn from the vast amount of research and evidence available to create a more equitable and impacting assessment system not just for ELs, but for all students.

District and School Leaders

Leadership at the district and school level need to include ELs into accountability systems in valid, appropriate, and meaningful ways that recognize the unique strengths and needs of this heterogeneous student population. Leaders need to accept the complexity of the matter and strive to create organizational cultures that promote child-centered cultural understanding. Schools and districts need to take a stand and advocate on behalf of children, families, best practices, and moral righteousness. Superintendents, principals, and other school leaders need to champion educational policies that support local language policies and practices (Menken, 2008). Policies that include multiple measures of student learning, like actual student work samples, as well as, classroom, school, and district assessments. Leaders need to push back on students, teachers, and schools being defined by the oversimplified notion of test scores from a single assessment.

In addition, district and school leaders need to reframe current narratives to emphasize the ability of ELs to learn and achieve. Current policies, systems, and rhetoric all fuel deficit-based approaches to this student population and community. Throughout all layers of the educational system, we need to adopt an asset-based and growth-minded perspective with respect to ELs and to the incredible promise they hold for our schools, communities, and state.

In the absence of state and federal policy, schools and districts need to develop clearly articulated language policies that support the goals and needs of their unique communities (Menken, 2008). The role of assessment, particularly standardized testing, needs to be clearly defined within this vision. Leaders need to support the creation of systems, supports, and educational programming aligned to best-practice research and not the output data of large-scale standardized assessment. This is not advocating for the abolition of assessment data, rather for the judicious and responsible handling of it. This information should inform only part of a much more comprehensive understanding of a child as a learner. Teacher observation, parent feedback, work samples, student self-reflection, classroom based formative and summative assessments, and a comprehensive understanding of a child's primary language proficiency should all be part of this detailed portrait of a learner.

School and district leaders need to use their knowledge, expertise, and positions of authority to promote educational opportunities that are equitable and impactful for all students, particularly culturally and linguistically diverse students that are overwhelmingly underserved by our current reality.

Teachers

As with curating the day-to-day educational opportunities for students, teachers play the most important role in negotiating the implementation of assessment and language policies. Teachers need to expand their professional learning with regard to assessment practices. More specifically, they must familiarize themselves with research regarding the reliability, validity, and equity of these practices for students who are culturally and linguistically diverse. Teachers need to have a common and clearly defined goal of working towards cultural validity. They must use their positions on the front lines of serving students to demand more equitable and valid assessment practices. Teachers must champion the cause of multiple measures to be used to assess and define student learning, and demand that the work produced in their classrooms is the best indicator of what a child knows and can do. Furthermore, teachers need to advocate for the judicious and responsible usage of assessment data at the classroom, grade, school, and district level. Teachers, particularly those with tenure, need to ask questions, push back, and help to reshape the oversimplified narratives in which assessment data is frequently discussed. Again, this is not advocating for the abolition of assessment data, rather for a more comprehensive handling of it as one part of a more robust student profile.

Above all, teachers need to know their students. They need to build the trust and relationships to really get to know each child, their family, their language, and their culture. Teachers need to move beyond culturally responsive as a buzzword of best practice, and actualize it in the concrete actions of teaching, learning, and daily interaction. Teachers need to plan instruction and assessment with each of their students in mind. They need to plan these experiences with the cultural preferences, characteristics, needs, and strengths of each child at the forefront, and be willing to flexibly adapt along the way (Basterra, 2011). Most importantly,

teachers need to create classroom communities that value the uniqueness and contributions of each child. Every child must see themselves as a learner, thinker, reader, writer, mathematician, artist, scientist, and more within our classrooms each and every day. Developing the relationships and environment for that to occur is the most important and greatest challenge facing a teacher in this work.

Recommendations for Future Research

As our schools continue to serve increasing numbers of culturally and linguistically diverse students and families, the issues related to the equity of assessment and education is paramount for all stakeholders. This study sought to examine the relationship between ELs' second language proficiency and their academic achievement on federally mandated standardized content-area assessments. Recommendations for further research based on the findings of this study include the following.

The current study could be built upon by exploring the challenges of ELs with the newer version of the PARCC assessment, the *Illinois Assessment of Readiness* (IAR), and comparing it to the current study to see if similar results were found. Though one less test for ELA/literacy and mathematics, the format, question types, and overarching blueprints are the same as PARCC. Furthermore, additional comparisons could be completed with other large-scale standardized assessments. Smarter Balanced is an assessment consortium similar to PARCC. The Smarter Balanced assessment was taken by over six million students across twelve states, the US Virgin Islands, and the Bureau of Indian education in 2017 (Smarter Balanced Assessment Consortium, 2019). Like PARCC, the number of states actively participating in the assessment consortium has dropped over the past several years, a decline from the beginning total of thirty states.

Unlike PARCC, the Smarter Balanced assessment is computer adaptive, with questions getting more or less challenging depending on student responses (Smarter Balanced Assessment Consortium, 2019). This change of test design would provide an interesting point of comparison with respect to the impact on the performance of ELs. Like PARCC, the Smarter Balanced assessment has accommodations for ELs that focus on time and setting. This commonality of ineffective accommodations provides a further point of comparability between these two assessments. The NWEA *Measures of Academic Progress* (MAP) assessment is another large-scale standardized assessment that could be used in a follow-up study to compare results. The MAP test is used by many Illinois districts, and throughout the country, as an interim standardized assessment to gather more timely information on student learning and progress relative to grade level content-area standards for reading and math. MAP is typically taken three times per year in the fall, winter, and spring. Like Smarter Balanced, MAP is a dynamic assessment that adjusts question difficulty depending on the responses of individual students (NWEA, 2019). Like these other large-scale standardized assessments the accommodations for ELs focus primarily on time and setting. Though MAP has the ability to read aloud both reading and math texts to all students, if enabled. Where PARCC, IAR, and Smarter Balanced are all criterion-referenced assessments, MAP is a norm-referenced assessment. Currently, MAP utilizes reference norms established from an "English-speaking, school age (grades K-11) population" of 10.2 million students in 2015 (NWEA, 2019). While this reference group makes the normed data subject to further scrutiny for ELs, raw RIT scores could be used as a reference point to run statistical tests like the ones in this study to compare the results. In addition to this, comparisons to other state level performance-based content-area assessments throughout the

country could be used to localize this study with respect to both the assessment type and student population.

The current study could also be expanded to include a larger sample size. There are several ways this study could be increased in scale. First would be to include all students, not just ELs, within a school or district. This would require the creation of additional language proficiency reference groups. The recommendations for such groups would be one featuring former ELs that have been reclassified and one for non-ELs. This variation of the study would provide more points of comparison and greater insight into the impact of English language proficiency on achievement as measured by standardized assessments. Furthermore, the fourth research question and seven regression models would offer tremendous insight into the significance of race/ethnicity and socioeconomic status within an entire school, district, or larger system. Another variation of this study, would be to apply the statistical tests to a broader set of data. In particular, the study could be replicated at a larger urban university or in educational systems that serve large numbers of culturally and linguistically diverse students. Furthermore, the study could be replicated with data from the state level. Though access to student data would be a challenge to consider. Even with a Freedom of Information Act request, race/ethnicity and socioeconomic variables are likely to not be accessible and thus not available for inclusion. That being said, a large-scale investigation into the relationship of ACCESS scores for ELs and scaled achievement scores on federally mandated content-area standardized assessments would yield data of significant weight for policymakers and other educational decision makers.

Further follow-up to this research could include a qualitative investigation of the students, families, teachers, and administrators involved in this study. Understanding these

various perspectives on assessment and its impact through the words of participants would be invaluable to further establishing a human connection with this research. Furthermore, it would allow the opportunity for those involved to tell their own story and to reframe the recurring narratives of this work.

Another area of ongoing study is the impact of various assessment accommodations. As has been highlighted throughout this work, the accommodations currently provided to ELs are ineffective and do not align with contemporary research in the field (Abedi, 2007, 2011; Abedi, Hofstetter, & Lord, 2004; Lara & Chia, 2011). Studies that further examine assessment accommodations like clear assessment language, reduced low-frequency vocabulary, reduced linguistic structures, personalized dictionaries, and shorter context-embedded texts that are meaningful to ELs would be of tremendous benefit towards the development of assessments that are more valid, reliable, and meaningful for all students.

Conclusion

Given the ongoing emphasis of standardized assessment as a driver for accountability and school improvement efforts both at the national and state level, there is particular relevance and significance to this study. Additionally, the current debate over immigration and political rhetoric in the United States represents a pivotal moment for addressing the educational opportunities and inequities for those children learning to speak English as a second language. There are profound societal implications to this work that extend well beyond performance on a standardized assessment. The equity in assessment and educational experience for ELs is critical to our work in schools of providing all students with the knowledge, understandings, strategies, and confidence to achieve success in a self-determined future.

If after all of the reading, research, and reflection, I was to offer a grand unifying theory on standardized assessment, it would simply be not to do it. Though, if we are going to, then we must at least attempt to make it as reliable, valid, and free from bias for all students as possible. As President Barack Obama said, "change will not come if we wait for some other person or some other time. We are the ones we've been waiting for. We are the change that we seek." We have the information, we have the understanding, and right now we have the opportunity to create an assessment system and responsibility structure that is more equitable and meaningful for ELs and all students. This dissertation is a summons to put our collective voices and actions towards creating the conditions that serve all children. The only thing that remains to be seen is now that we know better, do we have the courage to do better?

References

Abbott, M. L. (2011). *Understanding educational statistics using microsoft excel and spss*. Retrieved from https://ebookcentral.proquest.com

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, 8(3), 231-257.

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher,* 33(1), 4-14.

Abedi, J. (2006). Language issues in item-development. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of tst development* (pp. 377-398). Mahwah, NJ: Lawrence Erlbaum Associates.

Abedi, J. (2007). English language proficiency assessment and accountability under the NCLB Title III: An overview. In J. Abedi (Ed.), *English language proficiency in the nation: Current status and future practice.* Davis, CA: University of California, Davis.

Abedi, J. (2009). Computer testing as a form of accommodation for English language learners. *Educational Assessment,* 14, 195-211.

Abedi, J. (2011). Assessing English language learners:critical issues. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (49-71). New York, NY: Routledge.

Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP math performance and test accommodations: Interactions with student language background.* CSE Technical Report 536. Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing. Graduate School of Education and Information Studies (GSEIS), University of California.

Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1-28.

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of student language background on content-based performance: Analyses of extant data* (CSE Teach. Rep. No. 603). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16-26.

American Educational Research Association, National Council of Mathematics Educators, & American Psychological Association (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Amrein, A. & Berliner, D. (2002). *An analysis of some unintended and negative consequences of high-stakes testing.* Tempe: Education Policy Research Unit, Arizona State University.

Aronin, L., & Jessner, U. (2015). Understanding current multilingualism: What can the butterfly tell us? In U. Jessner & C. Kramsch (Eds.), *The multilingual challenge: Cross-disciplinary perspectives* (pp. 271-291). New York, NY: Mouton de Gruyter.

August, D., Carolo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice*, 20(1), 50-57.

August, D., & Hakuta, K. (Eds.) (1997). *Improving schooling for language-minority children: A research agenda.* Washington DC: National Academy Press.

Bailey, B. (2007). Heteroglossia and boundaries. In M. Heller (Ed.)*, Bilingualism: A social approach* (pp. 257–274). New York, NY: Palgrave.

Baker, C. & Wright, W. (2017). *Foundations of bilingual education and bilingualism.* Bristol, UK: Multilingual Matters.

Baker, E.L., Linn, R.L., Herman, J.L., & Koretz, D. (2002). *Standards for educational accountability systems* (Policy Brief 5, Winter). Los Angeles, CA: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Basterra, M.D., Solano-Flores, G., & Trumbull, E. (2011). *Cultural validity in assessment: Addressing linguistic and cultural diversity*. New York, NY: Routledge.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition.* Cambridge, MA: Cambridge University Press.

Blackledge, A., & Creese, A. (2010). *Multilingualism*. London, England: Continuum.

Bronkart, J.-P. (1995). Theories of action, speech, natural language, and discourse. In J.V. Wertsch, P. del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 75-91). Cambridge, MA: Cambridge University Press.

Bunch, G.C. (2014). The language of ideas and the language of display. Reconceptualizing 'academic language' in linguistically diverse classrooms. *International Multilingual Research Journal, 8,* 70-76.

Butler, F.A., & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations* (CSE Tech. Rep. No. 448). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

California State Dept. of Education, S. O. of B. B. E. (1984). *Schooling and Language Minority Students: A Theoretical Framework*. Retrieved from https://search-ebscohost-com.ezproxy.depaul.edu/login.aspx?direct=true&db=eric&AN=ED249773&site=ehost-live&scope=site

Canagarajah, S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *Modern Language Journal*, *95*(3), 401–417. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ943600&site=ehost-live&scope=site

Center for Applied Linguistics. (2018). *2017 annual report*. Retrieved from http://www.cal.org/who-we-are/annual-reports

Chaudron, C. (2001). Progress in language classroom research: Evidence from The Modern Language Journal, 1916-2000. *Modern Language Journal*, 85(1), 57–76. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ621535&site=ehost-live&scope=site

Chu, H.S. (1981). *Testing instruments for reading skills: English and Korean* (Grades 1-3). Fairfax, VA: Center for Applied Linguistics.

Cohen, J. (1988). *Statistical power analysis for behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, E., & Allen, A. (2013). Toward an Ideal Democracy: The Impact of Standardization Policies on the American Indian/Alaska Native Community and Language Revitalization Efforts. *Educational Policy*, *27*(5), 743–769. Retrieved from https://search-ebscohost-com.ezproxy.depaul.edu/login.aspx?direct=true&db=eric&AN=EJ1167396&site=ehost-live&scope=site

Cole, M. (1999). Culture-free versus culture-based measures of cognition. In R.J. Sternberg (Ed.), *The nature of cognition* (pp. 645-664). Cambridge, MA: MIT Press.

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, *21*(4), 617–41. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ364940&site=ehost-live&scope=site

Collier, V. (1989). How Long? A Synthesis of Research on Academic Achievement in a Second Language. TESOL Quarterly, 23(3), 509-531. doi:10.2307/3586923

Collier, V.P. (1995a). *Acquiring a second language for school.* Washington, DC: National Clearinghouse for English Language Acquisition.

Collier, V.P. (1995b). *Promoting academic success for ESL students: Understanding second language acquisition for school.* Elizabeth, NJ: New Jersey Teachers of English to Speakers of Other Languages-Bilingual Educators.

Collier, V.P. (1995c). Second language acquisition for school: Academic, cognitive, sociocultural, and linguistic processes. In J.E. Alatis, C.A. Straehle, M. Ronkin, B. Gallenberger (Eds.), *Georgetown University Round Table on Languages and Linguistics 1995* (pp. 311-327). Washington, DC: Georgetown University Press.

Collier, V.P., & Thomas, W.P. (2007). Predicting second language academic success in English using the Prism Model. In J. Cummins & C. Davison (Eds.), *International handbook of English language learning, Part 1* (pp. 333-348). New York, NY: Springer.

Collier, V. P., & Thomas, W. P. (2009). *Educating English learners for a transformed world*. Albuquerque, NM: Fuente Press.

Cook, H.G. (2012). How children learn in two (or more) languages. In E. Hamayan, & R.D. Freeman (2012), *English language learners at school: A guide for administrators* (pp. 39-42). Philadelphia, PA: Caslon.

Crawford, J. (1992a) *Language loyalties: A sourcebook on the official English controversy.* Chicago, IL: University of Chicago Press.

Crawford, J. (1992b) *Hold your tongue: Bilingualism and the politics of English-only.* Reading, MA: Addison-Wesley.

Crawford, J. (1999) *Bilingual education: History, politics, theory and practice.* Los Angeles, CA: Bilingual Education Services Inc.

Crawford, J. (2000) *At war with diversity: US language policy in an age of anxiety.* Clevedon, UK: Multilingual Matters.

Crawford, J. (2002) OBITUARY: The bilingual education act, 1968-2002. *Rethinking School Online,* 16 (4), 1-4. Retrieved from http://www.rethinking schools.org/archive/16_04/Bil164.shtml

Creswell, J. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. New York, NY: SAGE Publications Ltd.

Cummins, J. (1981). Age of arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 11*(2), 132-149.

Cummins, J. (1989). *A theoretical framework for bilingual special education*. *Exceptional Children* (56, 111–19). Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ399079&site=ehost-live&scope=site

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire. Bilingual Education and Bilingualism*. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED446442&site=ehost-live&scope=site

Cummins, J. (2009). Preface. In D. Fu (Ed.), *Writing between languages: How English language learners make the transition to fluency, Grades 4–12* (pp. ix–xii). Portsmouth, NH: Heinemann.

Cummins, J. (2012). How long does it take for an English language learner to become proficient in a second language. In E. Hamayan, & R.D. Freeman (2012), *English language learners at school: A guide for administrators* (pp. 37-39). Philadelphia, PA: Caslon.

De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7-21. http://dx.doi.org/10.1017/S1366728906002732

Donaldson, M. 1978. *Children's minds.* New York, NY: Norton.

Durán, R. (2011). Ensuring valid educational assessments for ELL students: Scores, score interpretation, and assessment uses. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (115-142). New York, NY: Routledge.

Ellis, R. (2012). *Language teaching research and language pedagogy.* Malden, MA: Wiley-Blackwell.

Escamilla, K. (2012). What role does the home language play in an English language learner's life/schooling? In E. Hamayan & R.D. Freeman (2012), *English language learners at school: A guide for administrators* (pp. 49-56). Philadelphia, PA: Caslon.

Evans, B. and Hornberger, N. (2005). No Child Left Behind: Repealing and unpeeling federal language education policy in the United States. *Language Policy* 4, 87-106.

Farr, B., & Trumbull, E. (1997). *Assessment alternatives for diverse classrooms.* Norwood, MA: Christopher-Gordon.

Ferguson, A.M. (1985). Language experience for problem solving in mathematics. *Reading Teacher*, 38, 504-507.

Fowler, F. (2000) *Policy studies for educational leaders: An introduction.* Upper Saddle River, NJ: Merrill.

Gándara, P. (2015). Charting the relationship of English learners and the ESEA: One step forward, two steps back. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 1*(3), 112-128. doi:10.7758/rsf.2015.1.3.06

Gándara, P., & Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives,* 11(36).

Garcia, E. (1994). *Understanding and meeting the challenge of student cultural diversity.* Boston, MA: Houghton Mifflin.

García, O. (2009). *Bilingual education in the 21st century: A global perspective.* Malden, MA and Oxford, England: Blackwell/Wiley.

García, O. (2013). Theorizing and enacting translanguaging for social justice. In A. Blackledge & A. Creese (Eds.), *Heteroglossia as practice and pedagogy* (pp. 199-216). New York, NY: Springer.

García, O., & Li, W. (2014). *Translanguaging: Implications for language, bilingualism and education.* Basingstoke, England: Palgrave Pivot.

García, G.E., & Pearson, P.D. (1991). The role of assessment in a diverse society. In E.H. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 253-278). New York, NY: Teachers College Press.

Gass, S. (1998). Apples and oranges: Or, why apples are not oranges and don't need to be. A response to Firth and Wagner. *Modern Language Journal*, 82, 83-90.

Genesee, F. (Ed.). (1994). *Educating second language children: The whole child, the whole curriculum, the whole community.* Cambridge, MA: Cambridge University Press.

González, J. (2002, Summer). Editor's introduction: Bilingual education and the federal role, if any….*Bilingual Research Journal, 26*(2), i-v.

Gonzàlez, V. (2005). Cultural, linguistic, and socioeconomic factors influencing monolingual and bilingual children's cognitive development. *NABE Review of Research and Practice, 3,* 67-104.

Gorter, D., & Cenoz, J. (2010, March). *Developing multilingualism in school contexts.* Paper presented at the American Association for Applied Linguistics Atlanta, GA.

Hafner, A.L. (2001, April). *Evaluating the impact of test accommodations on test scores of LEP students and non-LEP students.* Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Hakuta, K., Butler, Y. G., Witt, D., & California Univ., Santa Barbara. Linguistic Minority Research Inst. (2000). *How long does it take English learners to attain proficiency?* Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED443275&site=ehost-live&scope=site

Haladyna, T.M., & Downing, S.M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.

Hamayan, E., & Freeman, R. D. (2012). *English language learners at school: A guide for administrators*. Philadelphia, PA: Caslon.

Herdina, P.H., & Jessner, U. (2002). *A dynamic model of multilingualism-perspectives of changes in psycholinguistics.* Clevedon, UK: Multilingual Matters.

Heubert, J.P., & Hauser, R.M. (Eds.) (1999). *High stakes testing for tracking, promotion, and graduation.* Committee on Appropriate Test Use, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academy Press.

Ho, A.D. (2008) The problem with “proficiency”: Limitations of statistics and policy under No Child Left Behind. *Educational Researcher, 37*(6), 351-360.

Hopkins, M., Thompson, K. D., Linquanti, R., Hakuta, K., & August, D. (2013). Fully accounting for English learner performance: A key issue in ESEA reauthorization. *Educational Researcher*, *42*(2), 101–108. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ997079&site=ehost-live&scope=site

Illinois Report Card (n.d.). *State snapshot*. Retrieved from https://www.illinoisreportcard.com/State.aspx

Illinois State Board of Education. (2018). *2017 Annual report*. Retrieved from https://www.isbe.net/Pages/2017-Annual-Report.aspx

Illinois State Board of Education. (2019). *English learners forms and notifications.* Retrieved from https://www.isbe.net/Pages/English-Learners-Forms-and-Notifications.aspx

Illinois State Board of Education (2019). *Illinois assessment of readiness (IAR).* Retrieved from https://www.isbe.net/iar

Illinois State Board of Education. (2019, November 21). NEWS: State board of education wants to know what educators and parents think about state assessments [Press release]. Retrieved from https://www.isbe.net/Lists/News/NewsDisplay.aspx?ID=1298

Jessner, U., Allgäuer-Hackl, E., & Hofer, B. (2016). Emerging multilingual awareness in educational contexts: From theory to practice. *Canadian modern language review*, *72*(2), 157–182. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1126258&site=ehost-live&scope=site

Jorgensen, J. N. (2010). *Languaging: Nine years of poly-lingual development of young Turkish-Danish grade school students*. Copenhagen Studies in Bilingualism, the Køge Series, K15–K16. Copenhagen, Denmark: University of Copenhagen.

Kaestle, F.F. (2001) Federal aid to education since World War II: Purposes and politics. In *The future of the federal role in elementary and secondary education* (pp. 13-35). Washington, DC: Center on Education Policy. Retrieved from http://www.ctredpol.org/pubss/futurefederal_esa/future_fed_role_kaestle.htm.

King, L. (2017). *The impact of multilingualism on global education and language learning.* Cambridge, UK: Cambridge English Language Assessment.

Kirkhart, K.E. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice*, 16(1), 1-12.

Kloss, H. (1977) *The American bilingual tradition*. Rowley, MA: Newbury House.

Kopriva, R., & Council of Chief State School Officers, W. D. (2000). Ensuring Accuracy in Testing for English Language Learners. Retrieved from https://search-ebscohost-com.ezproxy.depaul.edu/login.aspx?direct=true&db=eric&AN=ED454703&site=ehost-live&scope=site

Kopriva, R., & Sexton, U. (2011). Using appropriate assessment processes: How to get accurate information about the academic knowledge and skills of English language learners. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (96-114). New York, NY: Routledge.

Koretz, D. (2008). *Measuring up: What educational testing really tells us.* Cambridge, MA: Harvard University Press.

Koretz, D. (2008). The pending reauthorization of NCLB: An opportunity to rethink the basic strategy. In G.L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 9-26). Thousand Oaks, CA: Corwin Press.

Kramsch, C. (2014). Teaching foreign languages in an era of globalization: Introduction. *The Modern Language Journal*, 98(1), 296-311. Retrieved from http://www.jstor.org.ezproxy.depaul.edu/stable/43651759

Kramsch, C., & Zhang, L. (2018). *The multilingual instructor.* Oxford, UK: Oxford University Press.

Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom*. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED230069&site=ehost-live&scope=site

Ladson-Billings, G. (1995). Making mathematics meaningful in multicultural contexts. In W.G. Secada, E. Fennema, & L.B. Adjian (Eds.), *New directions for equity in mathematics education* (pp. 126-145). Cambridge, MA: Cambridge University Press.

Lara, J., & Chia, M. (2011). Overview of current federal policies in assessment and accountability and their impact on ELL students. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (168-191). New York, NY: Routledge.

Linn, R.L., & Gronlund, N.E. (1995). *Measuring and assessment in teaching*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Larsen-Freeman, D. (2017). Complexity theory: The lessons continue. In L. Ortega & Z. H. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 11–50). Amsterdam: John Benjamins.

Larsen-Freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals*, *51*(1), 55–72. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1172876&site=ehost-live&scope=site

Larsen-Freeman, D., & Tedick, D. J. (2016). World language teaching. *Handbook of research on teaching* (pp. 1335–1387). Washington, DC: American Educational Research Association.

Lewis, G., Jones, B., & Baker, C. (2012a). Translanguaging: Developing its conceptualisation and contextualisation. *Educational Research and Evaluation*, 18(7), 655–670.

Leyba, C. (1978). *Longitudinal study, Title VII bilingual program, Santa Fe public schools, Santa Fe, New Mexico*. Los Angeles, CA, National Dissemination and Assessment Center, California State University.

Linn, R.L. (2005). Conflicting demands of no child left behind and state systems: Mixed messages about school performance. *Educational Policy Analysis Archives, 13(33),* Retrieved from http://epaa.asu.edu/aojs/article/view/138/264

Linn, R.L. (2008). Toward a more effective definition of adequate yearly progress. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 27-42). Thousand Oaks, CA: Corwin Press.

Lynch, K., Grummell, B., & Devine, D. (2015). *New managerialism in education: Commercialization, carelessness and gender*. Basingstoke: Palgrave Macmillan.

Lyons, S. & Dadey, N. (2017). *Considering English language proficiency within systems of educational accountability under the Every Students Succeeds Act*. Retrieved from https://www.latinopolicyforum.org/publications/reports/document/Considerations-for-ELP-indicator-in-ESSA_030817.pdf

Menken, K., National Clearinghouse for Bilingual Education, W. D., & George Washington Univ., Graduate School of Education and Human Development. (2000). *What are the critical issues in wide-scale assessment of English language learners?* Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED450595&site=ehost-live&scope=site

Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Clevedon, UK: Multilingual Matters.

Menken, K. (2009). Policy failures: No Child Left Behind and English language learners. In S. Groenke & A. Hatch (Eds.), *Critical pedagogy and teacher education in the neoliberal era: Small openings* (pp. 49-63). Berline, Germany: Springer.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher,* 23(2), 13-23.

Michael-Luna, S. & Canagarajah, S. (2007). Multilingual academic literacies. *Journal of Applied Linguistics, 4*(1), 55-77.

Migration Policy Institute Data Hub (2018). *State immigration data profiles: Language & education.* Retrieved from www.migrationpolicy.org/data/state-profiles/state/language/IL/US

Mitchell, Corey (2017). *Thousands of English-learners fall short on test of language skills.* Retrieved from https://www.edweek.org/ew/articles/2017/07/19/thousands-of-english-learners-fall-short-on-test.html

Moran, R. (1988). The politics of discretion: Federal intervention in bilingual education. *California Law Review* 76(6): 1249-351. Retrieved from http://scholarship.law.berkeley.edu/californialawreview/vol76.iss6/2/

Moyer, M. (2016). Translanguaging: Language, bilingualism and education. *Journal of Sociolinguistics*, *20*(1), 125–130. Retrieved from: https://doi-org.ezproxy.depaul.edu/10.1111/josl.12171

National Center for Education Statistics (2018). *English language learners in public schools*. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp

National School Lunch Program (NSLP) Fact Sheet. (2019, March 20). Retrieved December 15, 2019, from https://www.fns.usda.gov/nslp/nslp-fact-sheet

NWEA (2019). *MAP growth.* Retrieved from https://www.nwea.org/map-growth/

Olson, David R. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review*, XLVII, 3, 257-281.

Otsuji, E., & Pennycook, A. (2010). Metrolingualism: Fixity, fluidity and language in flux. *International Journal of Multilingualism*, 7(3), 240–254.

Ovando, C.J., Combs, M.C., & Collier, V.P. (2006). *Bilingual and ESL classrooms: Teaching in multicultural contexts*. New York, NY: McGraw-Hill.

PARCC Assessment (n.d.). *Test design.* Retrieved from https://parcc-assessment.org/test-design/

Pugliese, R., & Filice, S. (2012). Plurilingual communication: A polygot model for a polygot world. In C. Argondizzo (Ed.), *Creativity and innovation in language education* (pp. 85-110). Bern, Switzerland: Peter Lang.

Ricento, T. (1995) A brief history of language restrictionism in the United States. In S. Dicker, R. Jackson, T. Ricento and K. Romstedt (eds) *Official English? No!: TESOL's Recommendations for Countering the Official English Movement in the U.S.* Washington, DC: TESOL.

Riddle, W. (1999) Education for the disadvantaged: ESEA Title I reauthorization issues. *Congressional research service issue brief.* Washington, DC: The Library of Congress.

Rivera, C. (1984). Language proficiency and academic achievement. *Multilingual Matters* 10. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED240882&site=ehost-live&scope=site

Rivera, C., & Vincent, C. (1997). High school graduation testing: Policies and practices in the assessment of English language learners. *Educational Assessment,* 4(4), 335-355.

Roseberry, A., Warren, B., & Conant, F. (1992). *Appropriating scientific discourse: Findings from language minority classrooms* (Working paper 1-92). Cambridge, MA: TERC.

Rosier, P., and Marilyn, F. Bilingual education at Rock Point-some early results. *TESOL Quarterly*, X, 4, 379-388.

Ruiz, R. (1984). Orientations in language planning. *NABE Journal, 8* (2), 15-34.

Sanchez, C. (2017). *English Language Learners: How Your State Is Doing.* Retrieved from https://www.npr.org/sections/ed/2017/02/23/512451228/5-million-english-language-learners-a-vast-pool-of-talent-at-risk.

Shepard, L., Taylor, G., & Betebenner, D. (1998). *Inclusion of limited-English proficient students in Rhode Islands' grade 4 mathematics assessment performance assessment* (CSE Technical Report 486). Boulder, CO: Center for the Study of Evaluation, Standards, and Student Testing.

Smarter Balanced Assessment Consortium (2019). *What is smarter balanced?* Retrieved from: http://www.smarterbalanced.org/about/

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilities views of language in the testing of English language learners. *Educational Researcher,* 37(4), 189-199.

Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices: an introduction. In M. Basterra, E, Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (3-21). New York, NY: Routledge.

Solano-Flores, G., Lara, J., Sexton, U., & Navarrete, C. (2001). *Testing English language learners: A sampler of student responses to science and mathematics test items.* Washington, DC: Council of Chief State School Officers.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553-573.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher,* 32(2), 3-13.

Solano-Flores, G., Trumbull, E., & Nelson-Barber, S. (2002). Concurrent development of dual language assessments: An alternative to translating tests for linguistic minorities. *International Journal of Testing*, 2(2), 107-129.

Stapleton, P., & Shao, Q. (2018). Research in language teaching over two decades: A retrospective of the first 20 volumes of "language teaching research." *Language Teaching Research*, *22*(3), 350–369. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1179714&site=ehost-live&scope=site

Sugarman, J. & Geary, C. (2018). *English learners in Illinois: Demographics, outcomes, and state accountability policies*. Retrieved from https://www.migrationpolicy.org/programs/ell-information-center

Thurlow, M., Liu, K., Erickson, R., Spicuzza, R., & El Sawaf, H. (1996, August). *Accommodations for students with limited English proficiency: Analyses of guidelines from states with graduation exams* (Minnesota Rep. 6). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Trumbull, E., Basterra, M., & Solano-Flores, G. (2011). Reflections on the promise of cultural validity in assessment. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (277-287). New York, NY: Routledge.

Trumbull, E., & Koelsch, N. (2011). Language arts: Designing and using a reading assessment for learners transitioning to English-only instruction. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (195-217). New York, NY: Routledge.

Trumbull, E., & Solano-Flores, G. (2011). Addressing the language demands of mathematics assessments: Using a language framework and field research findings. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (218-253). New York, NY: Routledge.

Trumbull, E., & Solano-Flores, G. (2011). The role of language in assessment. In M. Basterra, E. Trumbull, G. Solano-Flores (Eds.), *Cultural validity in assessment* (22-45). New York, NY: Routledge.

U.S. Census Bureau (2011). *Profile of General Population and Housing Characteristics.* Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

U.S. Department of Education, Office of English Language Acquisition, National Clearinghouse for English Language Acquisition. February 2017. *Profiles of English learners (ELs)*. http://www.ncela.us/files/fast_facts/05-19-2017/ProfilesOfELs_FastFacts.pdf.

Valenzuela, A. (2002). High-stakes testing and U.S.-Mexican youth in Texas: The case for multiple compensatory criteria in assessment. *Harvard Journal of Hispanic Policy* 14, 97-116.

Valenzuela, A. (ed.) (2005). *Leaving children behind: how "Texas-style" accountability fails Latino youth.* New York, NY: State University of New York Press.

Velasco, P., & García, O. (2014). Translanguaging and the writing of bilingual learners. *Bilingual Research Journal*, *37*(1), 6–23. Retrieved from https://ezproxy.depaul.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1029173&site=ehost-live&scope=site

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education.* Buckingham, UK: Open University Press.

Wells, G. (1979). Describing children's linguistic development at home and at school. *British Educational Research Journal*, V, 75-89.

Whetzel, C. (2017). Making sure that schools measure Up. *Education Week,* 18-20. Retrieved from www.edweek.org/go/qc17.

WIDA (n.d.). *ACCESS tests.* Retrieved from https://wida.wisc.edu

Wiley, T. and Wright, W. (2004) Against the undertow: Language-minority education policy and politics in the 'age of accountability'. *Educational Policy* 18(1), 142-168.

Williams, C. (1996). Secondary education: Teaching in the bilingual situation. In C. Williams, G. Lewis, & C. Baker (Eds.), *The language policy: Taking stock* (pp. 193-211). Llangefni, Wales: Canolfan Astudiaethau Iaith.

Wright, W. E. (2010). Landmark court rulings regarding English language learners. Retrieved from http://www.colorincolorado.org/article/landmark-court-rulings-regarding-english-language-learners

Zong, J. & Batalova, J. (2015). *The limited English proficient population in the United States*. Retrieved from Migration Policy Institute website: https://www.migrationpolicy.org/article/limited-english-proficient-population-united-states

2012 English Language Development Standards. (n.d.). Retrieved December 3, 2018, from https://wida.wisc.edu/resources/2012-english-language-development-standards

LIST OF APPENDICES

Appendix A

Critical (Exclusion) Values for Pearson's Correlations Coefficient

One-Tailed	0.05	0.025	0.01	0.005	0.0025	0.0005
Two-Tailed	0.1	0.05	0.02	0.01	0.005	0.001
Degrees of Freedom						
1	0.988	0.997	0.9995	0.999877	0.9999692	0.99999877
2	0.9000	0.95	0.9800	0.99000	0.99500	0.99900
3	0.805	0.878	0.9343	0.9587	0.974	0.99114
4	0.729	0.811	0.882	0.9172	0.9417	0.9741
5	0.669	0.754	0.833	0.875	0.9056	0.9509
6	0.621	0.707	0.789	0.834	0.870	0.9249
7	0.582	0.666	0.750	0.798	0.836	0.898
8	0.549	0.632	0.715	0.765	0.805	0.872
9	0.521	0.602	0.685	0.735	0.776	0.847
10	0.497	0.576	0.658	0.708	0.750	0.823
11	0.476	0.553	0.634	0.684	0.726	0.801
12	0.457	0.532	0.612	0.661	0.703	0.780
13	0.441	0.514	0.592	0.641	0.683	0.760
14	0.426	0.497	0.574	0.623	0.664	0.742
15	0.412	0.482	0.558	0.606	0.647	0.725
16	0.400	0.468	0.543	0.590	0.631	0.708
17	0.389	0.456	0.529	0.575	0.616	0.693
18	0.378	0.444	0.516	0.561	0.602	0.679
19	0.369	0.433	0.503	0.549	0.589	0.665
20	0.360	0.423	0.492	0.537	0.576	0.652
25	0.323	0.381	0.445	0.487	0.524	0.597
30	0.296	0.349	0.409	0.449	0.484	0.554
35	0.275	0.325	0.381	0.418	0.452	0.519
40	0.257	0.304	0.358	0.393	0.425	0.490
45	0.243	0.288	0.338	0.372	0.403	0.465
50	0.231	0.273	0.322	0.354	0.384	0.443
60	0.211	0.25	0.295	0.325	0.352	0.408
70	0.195	0.232	0.274	0.302	0.327	0.380
80	0.183	0.217	0.257	0.283	0.307	0.357
90	0.173	0.205	0.242	0.267	0.290	0.338
100	0.164	0.195	0.230	0.254	0.276	0.321

Source: Perason, E.S., and Harley, H.O. *Biometrika Tables for Statisticans*, Volume I., 2nd ed. Cambridge, UK: Universtiy Press, 1962. Adapted from Abbott (2011).

Appendix B

Exclusion Values for the T Distribution

One-Tailed	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025
Two-Tailed	0.8	0.5	0.2	0.1	0.05	0.02	0.01	0.005
Degrees of Freedom								
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32
2	0.289	0.816	1.886	2.92	4.303	6.965	9.925	14.089
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029
8	0.262	0.706	1.397	1.86	2.306	2.896	3.355	3.833
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690
10	0.260	0.700	1.372	1.812	2.228	2.764	3.196	3.581
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428
13	0.259	0.694	1.350	1.771	2.160	2.65	3.012	3.372
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252
17	0.257	0.689	1.333	1.74	2.110	2.567	2.898	3.222
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860
infinite	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807

Source: Perason, E.S., and Harley, H.O. *Biometrika Tables for Statisticans*, Volume I., 2nd ed. Cambridge, UK: Universtiy Press, 1962. Adapted from Abbott (2011).